Soc

NAVAL RESEARCH LOGISTICS QUARTERLY

OFFICE OF NAVAL RESEARCH

CONTENTS

Proceedings of Decennial Logistics Research Conference Under the Joint Sponsorship of the Office of Naval Research and The George Washington University, 20-22 January 1960.

eface 293

TRODUCTORY REMARKS

e Need for a Strategy of Logistics Research 295
Nathan Brodsky

ENARY SESSION

e Accomplishments of Logistics Research 299
W. H. Marlow

esearch for Command Logistics 315
E. K. Scofield

e Aspects of Mathematics in Social Sciences 335
C. Tompkins

ANEL 1 - HARDWARE FOR LOGISTICS RESEARCH

ntrol of and Communication with Data-Handling Machines 357
William Orchard-Hays

e Reflection of Logistics in Electronic Computer Design 365
E. W. Cannon

Automatic Addressing Device 373
J. P. Fennell

e Marriage of On-Line Human Decision With Computer Programs 379
H. M. Teager

ANEL 2 - NUMERICAL DESCRIPTION OF THE ECONOMY

ments on Interindustry Economics 385
A. S. Manne

(Contents Continues on Inside Back Cover)

Los Angeles

VOL. 7, NO. 4 DECEMBER 1960

VEXOS P-1278

NAVAL RESEARCH LOGISTICS QUARTERLY

EDITORS

Rear Admiral H. E. Eccles, USN (Retired)
The George Washington University

O. Morgenstern
Princeton University

F. D. Rigby
Office of Naval Research

D. M. Gilford
Office of Naval Research

Commander H. D. Moore, SC, USN
Managing Editor
Office of Naval Research
Washington 25, D.C.

ASSOCIATE EDITORS

R. Bellman, RAND Corporation
J. C. Busby, Jr., Commander, SC, USN
J. S. Campbell, Technical Operations Inc.
W. W. Cooper, Carnegie Institute of Technology
J. G. Dean, Captain, SC, USN
G. Dyer, Vice Admiral, USN (Retired)
P. L. Folsom, Captain, USN
M. A. Geisler, RAND Corporation
A. J. Hoffman, General Electric Company
Harris P. Jones, Commander, SC, USN
S. Karlin, Stanford University
H. W. Kuhn, Princeton University
J. Isaderman, Office of Naval Research, Br. Office
W. H. Marlow, The George Washington University
R. E. McShane, Vice Admiral, USN (Retired)
W. F. Millson, Captain, SC, USN
M. I. Rosenberg, Captain, USN (Retired)
D. Rosenblatt, The George Washington University
H. A. Sachaklian, Colonel, USAF
E. K. Scofield, Captain, SC, USN
J. R. Simpson, Bureau of Supplies and Accounts
J. S. Skoczylas, Colonel, USMC
H. Solomon, The George Washington University
I. Stakgold, Northwestern University
E. D. Stanley, Captain, SC, USN
C. Stein, Jr., Captain, SC, USN (Retired)
R. M. Thrall, University of Michigan
C. B. Tompkins, University of California
J. D. Wilkes, Office of Naval Research

The Naval Research Logistics Quarterly is devoted to the dissemination of scientific information in logistics and will publish research and expository papers, including those in certain areas of mathematics, statistics, and economics, relevant to the over-all effort to improve the efficiency and effectiveness of logistics operations.

Information for Contributors is indicated on inside back cover.

The Naval Research Logistics Quarterly is published by the Office of Naval Research in the months of March, June, September, and December and can be purchased from the Superintendent of Documents, U.S. Government Printing Office, Washington 25, D.C. Subscription Price: $1.50 a year in the U.S. and Canada, $2.00 elsewhere; $0.50 for a single copy. Reprints of an individual article can be purchased in quantities of 100 or more. Requests for the purchase price of reprints of a particular article should be sent to the Superintendent of Documents, U.S. Government Printing Office, Washington 25, D.C.

The views and opinions expressed in this quarterly are those of the authors and not necessarily those of the Office of Naval Research.

Use of funds for printing this publication approved by the Director of the Bureau of the Budget 7 July 1958.

Permission has been granted to use the copyrighted material appearing in this publication.

PREFACE

This issue of the Naval Research Logistics Quarterly contains the majority of the papers presented at the Decennial Logistics Research Conference. The papers are arranged in the order that they were presented at the conference, which was jointly sponsored by the Office of Naval Research and The George Washington University and was held at the University on January 20, 21, and 22, 1960. This marked the tenth anniversary of the Logistics Research Project at the University and the sixth Logistics Conference. The papers are arranged in the order of their presentation.

The first Logistics conference, held in January 1950, was the first working conference on mathematical problems in logistics. It represented a twofold effort typically found in the development of a science:

1. An effort to state the major problems in pertinent detail and in quantitative form with no necessary hope of an easy solution, but with some hope of general implicit knowledge concerning the problems;

2. The statement of the solution of some problems similar to, but smaller than, problems which appear to be generated under the first effort.

The Decennial Conference was also directed towards these same efforts. The topics at the 1950 Conference included various aspects of the inventory problem, scheduling problems, and a variety of "mixed" problems. The conferences which followed were devoted to a variety of subject-areas in logistics including organization theory, producer logistics, usage data, etc.

During the past ten years, many organizations and individuals have provided significant contributions to the development of logistics as a science. More and more attention has been focused on an "empirical" approach to logistics problems. In view of this, it was decided that an appropriate theme of the Decennial Conference would be "Empirical Methods in Logistics Research." The term "empirical" as meant here is not to be contrasted with "theoretical." Briefly, it does mean the gathering of relevant data, the search for structure of the data, the formulation of suitable models, and the subsequent tests of hypotheses.

After a Plenary Session where principal papers were presented, the conference was divided by subject matter into seven panels. Panel 1 was devoted to hardware for logistics research. The principal discussion here was directed at "hardware" from the viewpoint of the logistics-researcher, rather than that of the "engineer" or "designer." Briefly, the papers and discussions were devoted to the logical requirements of hardware for research in logistics. Panel 2, "Numerical Description of the Economy," was directed toward the "empirical" aspects of this broad subject. The main consideration here was the assignment of measurements to variables which are necessary for problems similar to those found in industrial-mobilization planning. Panels 3a and 3b were devoted to simulation techniques in logistics research. The distinction between the two is that one was concerned with the simulation of functions or the elemental aspects of problems while the other was directed toward discussions of the simulation of systems which include a collection of functions. Another way of stating this distinction is that the former represents a "micro" approach and the latter represents a "macro" approach to problems. Panel 4 had as its subject "Utility and Military Worth". Most of this discussion was directed at examples of the determination and use of measures of "worth." Panel 5 included discussions of recent problems and techniques under the heading "Transportation and Scheduling Problems." "Inventory Control" was the subject of Panel 6. Here the basic consideration was a discussion of statistical problems and techniques.

* * *

INTRODUCTORY REMARKS

THE NEED FOR A STRATEGY OF LOGISTICS RESEARCH

Nathan Brodsky

Office of the Secretary of Defense
Supply and Logistics

The Office of the Secretary of Defense welcomes the opportunity to participate in this Decennial Logistics Research Conference. In so doing, we extend our congratulations to the Office of Naval Research and to the George Washington University for some outstanding pioneering in this area. It is a tribute to this group that we are able to assemble now so many researchers of high caliber who are devoting their talents to the solution of logistics problems.

I would like to take a few minutes to talk with you about our research program and the need for a strategy of logistics research. I represent a combination of interests. Foremost, I represent the logistics manager who is faced with tremendous problems of uncertainty and unpredictability which he cannot solve solely on the basis of experience. He has used research in the past, but in recent years he has felt a greater need for systematic and comprehensive analysis in dealing with broader and longer-range problems. Impressed with the increasing application of the scientific method in the solution of problems in the social and behavioral sciences, the manager has been turning to research to provide him with an objective basis for decisions.

I also represent the interest of the researcher, who has found that he can make a valuable contribution to the solution of major problems in logistics. As a representative of the researcher, my task is to bring him together with the manager, so that he might demonstrate the value of the scientific attitude in the solution of the manager's problems. I am interested in showing those responsible for logistics decisions that through disciplined observation and rigorous proof they may discover new facts and relationships which will give them new dimensions in decision making.

The growing interest and investment of resources in logistics research has prompted the Office of the Secretary of Defense and the military departments to seek means for integrating this effort and to increase the operational application of usable findings. Toward these ends, the program is being extended to include not only contract research but also studies performed "in-house." We seek to direct our logistics research resources into the most practical channels of inquiry and to gear the research toward solving problems of consequence. All of us are agreed that we must encourage the maximum practicable interchange of research information throughout the DOD, and avoid unnecessary duplication of effort.

In order to integrate the program and obtain the maximum results from our efforts, we must allocate our resources judiciously. The allocation must include funds for basic research as well as applied. Basic research is an important element in the management sciences as well as in the physical sciences. We must constantly replenish the well of basic knowledge which eventually we may draw upon for the solution of our problems.

It is my belief that the proper allocation of resources requires formulation of a strategy of research, because the individual tugs and pulls of researchers do not necessarily produce the most desirable results. Dr. Calkins, of Brookings Institution, spoke of this problem in relation to economic research, and what he had to say applies aptly to our situation.

> "Many (scholars) feel that if only society would give them ample resources they would advance knowledge in the best or indeed the only possible way by pursuing their own interests in research. Given unlimited talent, time, and funds, this is a congenial and proven approach. But we have, in fact, limited talent, limited funds and urgencies for which there is limited time. We have, in brief, an economic problem of allocating limited resources, and hence a strategy is needed for the deployment of those resources to achieve the maximum results."

This statement rings many bells for us. We, too, have limited talents, limited funds, and many urgencies. We must recognize, therefore, that those of us concerned with solving major logistics problems must establish priorities of research. We must make the best use of our research resources. As those who possess the scarce talents which management requires, you researchers must work closely with management to isolate major problem areas and to seek solutions to problems of consequence. Bright Wilson, in his book on scientific research, appropriately says: "Many scientists owe their greatness not to their skills in solving problems, but to their wisdom in choosing them."

The development of a strategy cannot be a one-way street. Researchers and managers must work together harmoniously and with full understanding. Semantic curtains must be lifted. Messiah-complexes must be abandoned.

Now, what does this mean in specific terms? In brief, it means that while our logistics-research studies have been singularly successful, they have not yet optimized the use of our limited resources. In large measure, I believe that we have been operating on problems at a lower level than that at which major logistics decisions should be made. We have often done research in areas where methodological convenience was paramount rather than directing our attention to determining what the real problems are.

Much of our research has been directed toward inventory problems. This is certainly important, and the results have been useful. But I wonder whether many of us have not been attracted to inventory theory and its application largely because this is an area which lends itself to mathematical formulation.

I am suggesting that the scope of our research might well be extended in order to achieve a better balance of problem areas investigated. The economics of quality control, for instance, is a virgin field. Procurement, with its various problems of pricing, contract administration, redetermination, relationship to technological advance, etc., is largely untouched. A system for logistics-readiness evaluation has yet to be formulated. The means for achieving a closer relationship between research and development and logistics have not been adequately explored. The problems in these areas have not yet been defined.

Perhaps we should go back to a very fundamental investigation—what is "logistics?" I, for one, am not satisfied with definitions such as "economics is what economists do" or "operations research is what operations-researchers do" or "logistics is what logisticians do." I think we have a real problem in defining the area, and certainly researchers are in a favorable position to meet this basic need. Current definitions in the military dictionaries do not serve as adequate tools for us. It seems to me that we are concerned with three interrelated aspects which must be included in a definition: the relationship of logistics to strategic planning; the relationship to the national economy; and the management of the resources within the military economy.

Not only must we define "logistics" but we must also describe a logistics system conceptually. I do not mean that we need a charting of a supply system such as BUSANDA's or the Quartermaster Corps but rather a conceptual breakdown which describes the parts of a logistics system and their relationships to other parts within the system and to related parts outside the system. Let me illustrate briefly. Standardization is traditionally a part of logistics. Yet many standardization functions are of a purely engineering nature. Much standardization action is performed by engineers outside of logistics control. This tends to

create difficulties both for research and development people and for logisticians, not only because of competition for the same resources, but also because of lack of clarity for responsibility. This lack is not due to an unwillingness to assume responsibility, but rather to the inability to define the problem and then to proceed to seek a solution. A conceptualization of the logistics system is a basic step toward resolving such a problem.

Another area for research relates to the logistics support of missiles. There are some studies under way, but much more is needed to bring into sharper focus the logistics problems of the missile era and the suggested means for avoiding them. These studies must be broad in scope as well as comprehensive. They must concern themselves with decisions which have to be made at high levels in the logistics system. They must recognize that many of our traditional concepts of requirements-determination, procurement, transportation, distribution, inventory-management, communications, and maintenance need re-examination. The impacts of new-hardware developments upon logistics must be considered; the logistics limitations on the use of new weapons must be made clear.

In summary, I have suggested some major problem areas that need fruitful research and analysis. I spoke of the need for developing a strategy of research which would seek to allocate our research resources to achieve the maximum results. This strategy must be worked out jointly by researchers and those responsible for management decisions. It cannot be done unilaterally. You researchers have a vital role to play both in focusing attention on problem areas of significance and in working with management to program research directed toward these areas. We must not be content with applying familiar techniques; we must seek to explore major areas, to direct our best resources to those areas, and to ensure full understanding of what we are doing.

In this scientific era, we tend to develop a wider gap between the scientist and the ordinary mortal. The sociologist used to refer to this as a "cultural lag," but I fear that it has greater implications for us today. Our scientific method has to bring meaning to mankind if the human specie is to survive. We must, therefore, strive for understanding between the scientist and the public. We must not succumb to the easy way of isolating ourselves from the stream of life.

Bertrand Russell warns that "if we develop cleverness without wisdom" we shall inevitably become extinct. Reduced to the level of logistics research, if we develop subtle solutions without understanding, we shall have failed in our responsibility.

* * *

PLENARY SESSION

SOME ACCOMPLISHMENTS OF LOGISTICS RESEARCH*

W. H. Marlow

Logistics Research Project
The George Washington University

This is a survey paper drawing from several instances of work by the Logistics Research Project over the past ten years. The selection and treatment of this material has been influenced somewhat by the Conference papers being presented by other Project Members; however, this mainly tends to reduce the amount of detail reported for those areas. It is also true that the present paper tends in the direction of a historical record; empirical methods of logistics research are stressed but not to the exclusion of material on the origins of the problems attacked, the course of their formulation as research problems, or the practical application of the results. The principal subjects treated are: operational logistics planning, industrial mobilization, usage data collection and analysis, the Allowance List Test Program, logistics data processing, and some additional instances of logistics research. The rather ambitious aim of the paper is to present an account which will do justice to the many individuals who have been responsible for these accomplishments of logistics research.

INTRODUCTION

This is a survey paper drawing from several instances of work by the Logistics Research Project over the past ten years. The selection and treatment of material has been influenced somewhat by the Conference papers being presented by other Project Members; however, this mainly tends to reduce the amount of detail reported for certain of the areas. It is also true that the present paper tends in the direction of a historical record; empirical methods of logistics research are stressed but not to the exclusion of material on the origins of the problems attacked, the course of their formulation as research problems, or the practical application of the results.

In view of the long and continuing activity of the Project in basic research quite removed from the subject area of the present conference, it seems appropriate to start this summary of accomplishments by citing several papers sponsored by the Project which the reader may wish to consult: Fennell and Oshiro [5], Gourary [6, 7], Hannan [14], Heller and Tompkins [15], Isbell [19, 20], Isbell and Marlow [21, 22], and Kelley [23].

The principal subjects treated in the present account, which, as noted above, stresses empirical methods of logistics research, are the following: operational logistics planning, industrial mobilization, usage data collection and analysis, the Allowance List Test Program, and logistics data processing. The rather ambitious aim of the paper is to

*Research was supported by the Office of Naval Research.

present an account which will do justice to the many individuals who have been responsible for these accomplishments of logistics research.

OPERATIONAL LOGISTICS PLANNING

One of the first problems attacked by the Project concerned the determination of aggregated resupply requirements for support of naval forces. The initial approach was given in 1950 by C. B. Tompkins [33]. A research program based upon empirical methods was carried out at the Project by Mrs. Mina Haskind Gourary and others until the resultant methodology and data were incorporated within the official planning system for the Navy.

As the first step, variables were specified upon which usages would be based. Due attention was given to the connections between these variables, the consumptions, and the plan of action or the estimate of the situation. Resulting from these considerations were twelve independent potentials to use commodities if the commander so directed: consumptions of different types of fuel measured in units of average consumption rates under specified standard operating conditions; consumptions of ammunition against air, surface, subsurface targets measured in units of potential rates of fire; and consumptions of food, men, general stores, spare parts, and deck cargo, each measured in appropriate units.

In this scheme, which became "The Three Factor Method of Logistics Planning," requirements are computed as the product of three numbers. The first of these, the potential, has come to be called the activity factor since it is a measure of activity upon which consumption of material depends. The Project developed a set of these factors for naval vessels in 1951, for naval aircraft in 1952, and for Advance Base Functional Components in 1955. The sets of activity levels were obtained from naval sources and represented aggregated usage data. In fact, when these factors are multiplied by appropriate second factors, (the conversion factors), usages under standard conditions result. The usages are expressed in broad shipping units such as thousands of barrels, measurement tons, short tons, and long tons. The standard conditions are numerical descriptions of the type of activity from which these same usages would result. In all, from the twelve activity levels a total of about seventeen different categories of standard shipping requirements can be computed. Comparing the standard operating conditions with those being planned leads to determination of the judgment factor. That is, the third factor is a number stating the extent to which the commander intended to exploit a potential to consume.

Incorporation of this scheme and these data into the official Navy planning system started in 1953. This was partly in consequence of experience by the U.S. Naval War College in using this material for student determination of requirements in operations problems. Also, in 1952 the Project had collaborated with the Atlantic Fleet in order to calculate logistics requirements for a Naval Code Logistics Plan by the three factor method. Subsequent similar calculations have been made by the Navy on a regular basis.

In summary, the accomplishment of these planning factors was the provision of an effective means for determining resupply requirements at the major task force, fleet, or theater level. Through inclusion of means for introducing (and recording) quantitative judgment, not only were improved estimates of broad logistical requirements made possible but effective communications between planners were simplified.

The planning factors of the three-factor method have had utility for subsequent research at the Project in the area of operational naval logistics planning. As particularly noteworthy among these, there is the concept of mobile logistics support developed during 1956 by Captain

Ralph B. Hunt, USN (Ret.)[16]. This particular concept evolved as the end product of extensive empirical work. Furthermore, as will be described, this research itself led into a vast area for further empirical work in transportation scheduling which has been investigated since mid-1957 by Captain Hunt and E. F. Rosholdt.

The study which led Captain Hunt to the resupply concept had as its objective the investigation of implications of various methods of providing logistics support for a carrier striking force. Work began by fixing a representative force to be supported, developing a concept of operations for the force, and then devising a corresponding replenishment plan. All of these aspects were of course then fixed in order that there would be a common set of "pure logistical requirements." These latter were resupply requirements in broad commodity aggregates such as were described above in the account of the "Three-Factor Method;" they had to be provided by underway replenishment to the carrier striking force. The remainder of the investigation concerned how the underway replenishment group might itself be replenished (as, for example, via a base or directly from the US). Another way of describing this is to say that the resupply pipeline was to be studied.

Various methods for logistics support were tested for different theaters of operation such as the Atlantic, Pacific, and Indian Oceans. The testing by empirical methods was comparable with the planning which would have to be carried out by the various commands and staffs responsible for such logistics support operations. Far from being carried out in a vacuum, testing (calculation) was accomplished with actual Navy planning data or else with data which could be adjusted to fit Navy doctrine. One of the most novel features of this study was the comprehensive coverage of the various integral logistics aspects within one broad scheme. That is, the many detailed logistics considerations which were encompassed in this study were unusually complete and responsive to the many practical aspects of the military problem.

The type of logistics support which turned out to be most advantageous involved direct turnaround of the replenishment ships from the USA. Such favoring of direct support was a consequence, first of the greatly increased costs of alternative replenishment via an overseas base, and second of the amount of flexibility possible for the carrier striking force under direct support from the USA.

In terms of the critical ships involved (the Navy ships capable of underway transfers), this study brought out the dependence on the capacity of the replenishing ships rather than the distance from the resupply base. That is, the numbers of ships in the resupply pipeline were relatively insensitive to the distance between the resupply base and the replenishment area. For replenishment directly from the USA to the Atlantic, Pacific, and Indian Ocean areas with one-way distances in miles respectively 3200, 5400, 8600; refrigerator ships (AF's) required were 4, 4, 5, respectively. The above numbers of ships were the totals in the pipeline, and in each case they were to be reloaded only in the USA. Similar independence of distance prevailed for the numbers of other types of vessels, for example, the numbers of Navy oilers (AO's) turned out to be exactly the same for the three areas. There was some difference between theaters in petroleum requirements but only for the commercial type tankers which were required to replenish the oilers overseas. What prevailed here were the essentially complete utilization of the petroleum products on board plus the limited (relative to requirements) capacities of standard tankers. Where the ratios for standard tankers in the three theaters were near 3:15:13, triple-capacity supertankers would suffice in the ratios 3:4:9.

Operating through one overseas base with its required double handling of cargos increased the numbers of ships in the pipeline from 35 to 50 percent. The same percentages applied to the value of the pipeline plus extra costs directly attributable to support of the given carrier striking force. In fact, 65 percent of the total personnel ashore and afloat engaged in this resupply support could be eliminated by direct support from the USA. It should be noted that the above was strictly limited to that portion of the base concerned with the replenishment, so that only this one function of the base is discussed in the above.

The cited reductions in the numbers of ships in the pipeline are indications of the spread of initial "costs" for a continuing system of logistics support. To them must be added differences in recurring costs as, for example, the continuing cost of added base personnel. However, the principal war-time problem here, as in any extensive transportation area, is to set in motion the initial resupply pipeline. Increases in consumption rates for war-time plus additional obvious complications from any type of war make it difficult indeed. The great reductions possible by direct support from the USA caused this study to be of considerable practical significance.

The importance of the size of the pipeline plus the manner in which it varied sharply with capacities of the ships led to close investigation of general means for determining its "best" size. Corresponding to a time-phased schedule of required deliveries during each time period at specified consumption points, how can one determine the total number of ships required? This total number is of course determined after individual times per ship are assigned for each of the elements of the turnaround: loading time, time required in transit to delivery point, unloading time, and time for return voyage. One ship cycle, or turnaround, then is the sum of four component times. At any point in time, the pipeline size is the total number of ships engaged in turnaround.

How does one determine a "best" set of sailing schedules to meet stated delivery requirements? Captain Hunt and E. F. Rosholdt have performed extensive empirical research in this area which has yielded methodology permitting striking reductions in numbers of ships required [18]. Their methodology resulted from development of basic relationships in scheduling ocean transportation. Their approach involved graphical simulations of the scheduling problems, deriving computational algorithms from these and refining the techniques during the course of their application to actual problems. (For over two years members of the Project have been collaborating with various parts of the Defense Department in computing estimated shipping requirements consequent to instances of military plans.) In collaboration with J. P. Fennell, their scheduling work has been extended to application of the electronic Logistics Computer at the Project for determination of schedules and, of course, for further numerical simulation with the objective of improving the methodology.

What would be a precise characterization of the sense in which these sailing schedules are "best"? In the first place, it has to date been necessary to restrict attention to minimization over a finite set of schedules; this has been a consequence of constraints imposed by the actual military problems being attacked as part of the research. Indeed, one essential feature is that the operational characteristics required are quite specific and precise; for example, a "best" schedule in the sense of "minimizing" the pipeline which was obtained by always having the entire month's deliveries arriving on the last day of the month would not necessarily be "best" in the context of the military problem. What the Hunt-Rosholdt scheduling techniques yield are schedules which satisfy the military constraints and which are "smooth" in that they

minimize over time the maximum number of ships in the pipeline. Actually, they achieve "smoothness" in the sense that a successive "minimax" is obtained for each time period for which attention has been given to requirements of such other time periods as are affected because of the turnaround time. For example, if the turnaround time is more than one month, the sailings for a particular month are of course in response to requirements of future months. Their schedules are then those which produce minimum numbers of ships in the pipeline in the above sense.

It seems particularly worthwhile to note that these scheduling methods are applicable under widely varying planning situations. While reductions in the numbers of required ships are most striking in long-term schedules, application of the methodology to short-range (e.g., limited war) situations is not only desirable for minimizing ship requirements, but produces many side products useful for integrating the many logistics aspects which must be provided. For example, the methodology has been extended to improve scheduling of escorted convoys. What was done was to provide means for interlocking the schedules of the convoys and the escort groups to minimize requirements for the escorts, which are generally the limiting factor. It turned out that vast improvements over arbitrary or even "straight-forward" schedules are possible if the turnaround times for both the convoyed ships and the escort groups are multiples of the sailing intervals for the convoys.

As another accomplishment of this empirical work, methods have been devised for computing replacement vessels required because of repairs on the ships. Details are given in [18], but it can be remarked in summary that means were devised for realistic planning to replace application of percentage factors across the board. In solving some actual problems these latter were shown to be quite unacceptable. As a by-product of the sailing schedules which were being generated, it was possible to formulate a repair policy and then interpret it in terms of making use of ships leaving shipyards to replace ships entering for repairs. Furthermore, this policy was capable of being applied only to ships which are actual candidates in view of their location at a particular point in time.

Finally, the fundamental planning units for this scheduling work are notional ships which are standard reference ships. Notional ships amount to planning units for expressing shipping requirements. The present state of the art in the military establishment also dictates that a notional distance be fixed for each type. Heretofore, this was frequently a cause of unrealistic requirements being generated. Based on their work, Captain Hunt and Mr. Rosholdt have devised a more useful and more realistic type of notional ship, results of which can be translated into particular actual type vessels or into the standard military planning units [17].

In 1956 and again in 1957 the Project and the Naval War College in Newport, Rhode Island, were linked by a leased teletype line in order that the Logistics Computer might be employed rapidly to produce logistics data. These data were required for the Naval War College Strategic Planning Study and Strategic War Game and consisted of the following:

1. Build-up of military capabilities and availabilities following M-Day in intervals of three months for personnel, military production, resupply, POL, and transportation.

2. Military requirements for comparison with the above based on force tabs and planning factors, both strategic and logistic.

3. Net residual capabilities (or requirements) based on various types of nuclear attacks.

The first two parts provided bases for evaluating alternative courses of action in the pregame planning. The third enabled the interactions of the game to be assessed.

The use of the Logistics Computer provided the War College with a credible environment for student instruction wherein rapid recalculations corresponding to changing situations were possible. This work provided the Project with valuable research material in the following respects:

1. The model of global all-service war from the viewpoint of high command including aspects of both producer and consumer logistics provided means for measuring the contribution of logistics to success in war.

2. Experience was gained in simulating data-processing aspects of logistics similar to that likely to be encountered in a future war. This included remote-control aspects as well as heavy teletype traffic. Furthermore, additional insight into computer requirements were obtained [26].

The construction of this model involved extensive data gathering and empirical generation of numerical functions representing the many quantities required for this broad simulation of warfare. Principal innovators in this work were Captain Parker L. Folsom, USN, and George Stephenson. While it is difficult to evaluate the accomplishments other than those involving hardware [26], it has turned out that these studies have provided the basis for subsequent Navy work on modern command control centers. This came about because the essence of the results contributed experience and techniques in the correlation of strategy, tactics, and logistics by means of scientific techniques and automatic computing machinery. From a practical point of view, the contributions were improvements and extensions of classical military processes, namely: the estimation of situation and decision, the preparation and formulation of plans, and the supervision of planned action.

Before taking up the next topic, mention should be made of the accomplishments of Rear Admiral Henry E. Eccles, USN (Ret.), in the area of treating logistics from the point of view of command. His concern has been with some of the most fundamental aspects of logistics as a part of military science. His book, "Logistics in the National Defense" [3], should be required reading for all who are concerned with national defense, particularly those who are involved with research in logistics.

INDUSTRIAL MOBILIZATION

Since its first year of existence the Project has been concerned with problems of mobilization planning, chiefly from the viewpoint of eventual matching requirements with resources on at least the Navy bureau level. Considerable empirical work was done during 1951-1952 in order to obtain a model of the shipyard industry for integration into a quantitative description of the national economy which was prepared under the over-all supervision of the Bureau of the Budget. Specifically, the Project cooperated with the Air Force group under Mr. Marshall Wood and with the Munitions Board in connection with their feasibility test, which became known as "The Emergency Model" and which was based on methods of interindustry economics. The economy was aggregated into 192 industries of the Standard Industrial Classification (SIC), with the feasibility test as follows: given some stipulated rate of deliveries of items and materials to the military services, what is the total level of activity required for each of the industries?

The purpose of the Project's efforts was to derive a bill of goods in terms of the above 192 sectors for the shipyard industry on a quarterly basis for the years 1951 through 1954. Engineering methods were employed to determine inputs to the bill of goods as well as to determine lead times. Taking this approach, rather than employing accounting data, led to a particularly satisfactory model of the shipyard industry. For one thing, use was made of a

prototype concept, which consisted of selecting representative vessels for which data were available in some form and identifying each vessel in the construction program with its most comparable prototype. Two principal advantages of using this concept were that it permitted use of available data and it resulted in handling a much smaller volume of data than would otherwise have been necessary. From data collected by the Project from both private and government sources, a time-phased list of direct material requirements and an "overhead bill of goods," each in appropriate dollar units, were derived and allocated to the SIC categories. Given the final "bill of goods," the use of an input-output technique was used to estimate or forecast total requirements; i.e., the input requirements placed on industry directly by the ship program (bill of goods) plus those requirements generated throughout all industries in the economy because of the initial ship program requirements (the intermediate products).

For purposes of the present summary, attention will be limited to the accomplishments of the empirical research leading to the description of the ship-construction, ship-conversion, and ship-repair "industry." As has been indicated above, an engineering approach was taken rather than aggregate dollar values of activity being stipulated for the period (i.e., budget) and this total being broken down into requirements for the various industries, assuming some product mix. Some of the practical considerations which motivated constructing the bills of goods empirically by summing materials required for individual ships rather than taking the last mentioned approach were as follows:

1. The relatively long production period for a ship and the necessary orderly scheduling of material to be put in place;
2. The difference in the building periods between different types of ships;
3. The changing requirements on the economy due to changes over time in the types of ships being constructed.

The chosen approach with its useful production functions leading to good final demand was all the more satisfactory in view of empirical evidence which suggests that the problem of estimating the "capacity" of even one shipyard is approachable only through adequate statements of specified requirements. In other words, final demand seems to be the only suitable starting point for questions of such feasibility testing. As an example of some additional insight which was gained by the empirical work, it was found that classical methods of labor productivity analysis could contribute only to those portions of the shipyard industry which were concerned with "assembly" rather than "manufacture." Key results were the derivation of proper concepts for, and measures of, the output from the shipyard industry which were capable of being reconciled with a meaningful concept of capacity.

As a further accomplishment, the model was arranged for automatic computation so that it was possible to study alternatives in the sense of changing the building program. Calculations for such changes were easily carried out and provided a facility for varying programs in order to approach a feasible military plan. Similarly, effects of the program mix over time were susceptible to analysis as, for example, the effects of constant industry coefficients in the input-output matrix.

The above is a very brief indication of the contribution of the Project in one series of studies (see [28]). Although other instances exist (notably work on the Controlled Materials Plan approach to mobilization planning [34]), the present report will end by describing more recent work at the Project by Stedman B. Noble. Mr. Noble's work has used directed graphs and their equivalence to the existence of coefficients in an input-output matrix. The problems

studied have been the control of errors of aggregation, and both analytic and empirical measures of decomposition in these matrices. From this, significant spurious relationships resulting from aggregation in the input-output model can be eliminated. In his present work he is using directed graphs to reexamine and integrate methods of data collection, computation, and model formulation. The result is a synthesis where each of these methods is made more manageable than they are for the input-output model. His work shows that most industries of the economy have flows that go mainly to a limited part of the economy. Some of the flows that do occur, however, must be studied in considerable detail if an accurate model is to be obtained. His work shows the validity, in a general sense, of much of current mobilization planning. By the use of mathematical methods and data from the Census of Manufactures (which data form the basis of the 450 x 450 table of the Bureau of Labor Statistics), his work supplements current planning by showing how to make the existing procedures more formal and giving methods for using a basic source of data. This then is a brief account of Mr. Noble's work, which has included extensive empirical research on problems of industrial mobilization planning and feasibility testing. He gives a comprehensive account of his recent work in [27].

USAGE DATA COLLECTION AND ANALYSIS

The most fundamental aspect of empirical research by the Project in the area of line-item logistics planning has been the investigation of patterns of demands for specific items by, or on account of, specific consumers. From its beginning, the Project has had as one of its problem areas that of how best to set stock levels to provide for uncertainly known future usages. Basic considerations have been described by Captain Hamilton and Vice Admiral McShane in [13] as follows:

> "When C. B. Tompkins was Principal Investigator of the Logistics Research Project, he conceived the idea that if a sample of naval ships could be described in detail in terms of the materials which were used in support of the operation, maintenance, repair and alteration of these ships, basic logistics knowledge could be gleaned from proper analysis of these details. In fact, he viewed such a description as a necessary first step for laboratory work in "experimental logistics.
>
> "A reasonably thorough search was made through all probable naval sources for descriptive details of this kind and none could be discovered. Much data of these kinds are generated during regular supply and support procedures and, in general, the raw data are retained for two years and then destroyed. No evidence could be found that any of these raw data had been reduced and aggregated or consolidated in such a way as to be analyzable. The historical data readily available were in terms of money value which was adequate for naval fiscal and administrative purposes but lacked information as to the specific character of the material which had been used.
>
> "The problem thus presented was either give up the idea of getting hold of the material description desired or, alternatively, develop a plan to collect the raw data and reduce them to analyzable shape. Development of a plan for the latter was approved by the Project's contract sponsors. Beginning in late 1951 work was started on making up a plan. The collection program began on a trial basis in the summer of 1952."

Such was the background of the Korean Data Collection Program which resulted in the collection of five years of usage for a representative fleet sample of 65 ships and an additional year's usage for 27 of these ships. With regard to coverage of materials consumed, all supply documents were collected which represented usage of material by, or on account of, a sample ship excepting only the following categories: provisions, medical and dental supplies, material for resale such as clothing and small stores and ship's stores, training devices, and ship's fuel. In other words, only materials strictly for the satisfaction of human requirements and fuel were omitted. Those that were covered were: general stores (Navy Cognizances A, G, W); repair parts (Cognizances H, N, P, R, Z); and major components (Cognizances F, J, S, V). Further background is implicit in the following discussion of the unique features of this data-collection program.

First, the usages were identified with specific ships. This meant that measurements were being made on the demands which had to be satisfied to support specific ships. In contrast, most available data were (and are) based on issues, i.e., on material flowing through certain points in the system.

Second, as noted above, coverage cut across all cognizance lines. Only by such means was it possible to obtain a comprehensive coverage of the impact of these ships' demands.

Third, all data sources were covered, not merely the sample ship's use. A total of 119 different data sources in addition to the sample ships themselves became involved: tenders, repair ships, naval shipyards, other naval shore ship-repair facilities, and a large number of private ship-repair plants.

Fourth, no special questionnaires were used. Complete reliance was placed on existing Navy forms of various types which, in particular circumstances, could represent the last piece of paper generated in the Navy Supply System by an item before it passes out of the system for application.

Fifth, the period of collection was of sufficient length to permit valid statistical observations. Similar to this is the fact that there was good coverage over types of ships, particularly for submarines (12) and destroyers (9).

Finally, it seemed to be a unique feature of this data-collection program that 100 percent cooperation was received from the many activities which became involved. This might be considered all the more surprising in view of the fact that, from start to finish, the data collection was undertaken as a research venture—on the one hand, it was monitored continuously by the Project (Captain Hamilton and Mrs. Bettie Taggart) and never really became routine; on the other hand, there always existed at the Project very limited resources for coping with the staggering amount of data being received. Limited resources at the Project had a more acute effect on the data-reduction phase, which phase itself could be used to illustrate research problems requiring empirical work. However, in the present paper, it will suffice to note that a priority system was set stressing early reduction to a form suitable for computer input data on submarines and destroyers plus a few larger types.

Before empirical work with these data are considered, three summary accomplishments of the research which went into usage data collection and reduction should be noted (see [12]). First, preservation of the raw data was accomplished. Second, a great deal of information was accumulated about naval processes which has broadened the base of understanding both for research and for related operating problems including even other data-collection ventures by the Navy. Finally, the reduced data themselves have been made available to other research activities and to operating naval agencies.

In late 1956 a joint study in the area of requirements determination was initiated by the Project and the Bureau of Supplies and Accounts. The first report [1] dealt with analysis of the behavior of repair parts usage by all sources for twelve submarines over a four-year period. Extensive computer processing of the usage data produced tables of cumulative percentages of number of demands for individual items by vessel, total quantities demanded for individual items by vessel, numbers of items, number of demands, and so on. An important point is that in this analysis (and in the subsequent studies to be described) all of the usage data for the cognizance classes being studied were used—each piece of data found its way into the various summary tabulations. In other words, this was not an instance of taking a few selected items and then producing a few listings for manual perusal. Summary conclusions in [1] were as follows:

First, the strong similarity of the behavior of the data observed for the twelve submarines indicated that the demand data actually represented material ordered for end use. That is, the data were judged to represent usage rather than individual whims and caprices. The similarity actually extended to total numbers of demand documents for each ship by each source, the range of stock numbers used, the quantities used per stock number, and the total quantities demanded. Unfortunately, there was practically no commonality of usage between individual submarines (even sister boats of which there were three pairs) in terms of actual items. Further than this, on any one submarine in different time periods, demands were for different items. In terms of any one of the submarines, and covering all sources of usage for four years, about 75 percent of the items which could have been used (the installed-parts population) had usage zero. About 18.5 percent of the wearable parts showed one demand and only 7.5 percent were used more than once.

Second, the observation of low movement coupled with a varying range of movement for successive time periods pointed to the conclusion that shipboard stocking policies should give greater emphasis to the range of items carried rather than to their depth.

Third, given this evidence of an extremely low and variable demand pattern for an individual ship, it would appear that standard inventory control formulations do not apply to the problem of stocking repair parts on account of specific demands of consumers.

The above conclusions underscored the necessity of developing and testing new techniques for planning problems which require the determination of stock levels on an individual line item basis. Subsequent analysis of usage patterns [31] for five years of submarine data on additional categories of material added further evidence to support the above conclusions. Similar analysis producing quite similar usage patterns was carried out for two sister destroyers [8]. For example, in four years of usages for ships parts (Cognizance H), out of 2894 distinct stock numbers demanded by the one destroyer and 2578 for the other, only 746 items were common to both ships. Further empirical studies were performed on a heavy cruiser [9] covering ships parts and electronics. With the notable exception of electronic tubes which have relatively high usage rates, the demand statistics reflected low, sporadic usages with highly similar magnitudes for the different types of ships.

As a final note, reference should be made to one other study [10] concerning material usage by and on account of the submarine TIRU. Many facts have been developed empirically from "transaction-count" data covering more than eight years. In addition, and this was the main objective, an evaluation of the "completeness" of the collected data was carried out

which was necessary for further planned research as well as for additional evidence to support earlier empirical findings (see also [11]).

THE ALLOWANCE LIST TEST PROGRAM

It was noted above that empirical results on usage patterns by Navy ships underscored the necessity of developing and testing new techniques for planning problems which require the determination of line item requirements. Actually, considerably more than this has been revealed: not only have new techniques been shown to be needed but earlier problem formulations have had to be revised. Another way of expressing this is to observe that the empirical data have led to results concerning "what to solve" as well as "how." Certainly this was expected to be a result of the work in view of classical instances of research. Perhaps what could be described as surprising is the degree to which the data have assisted in deriving quantitative formulations which possess the dual attributes of being directed at the right problems and of being such to permit quantitative solutions. In order to develop the above summary remarks, it will be necessary to describe further work under the Allowance List Test Program.

The Allowance List Test Program, a joint effort by the Project and the Bureau of Supplies and Accounts evolved rapidly from the collaboration cited above in connection with the analysis of usage data for twelve submarines [1]. By an allowance list is meant a specification of range and depth for "repair parts" to be carried on board a vessel for use by the ship's force in direct support of installed components. This being the case, it is clear that in a general sense the allowance list problem can be viewed as "maximizing" the endurance of a ship, i.e., the capability for sustained operation independent of logistical support other than by underway replenishment or other "routine" ways which do not interfere with the ship's ability to carry out an assigned mission. For definiteness, the following discussion will be limited to submarines. (This also corresponds to the fact that initial efforts in the Allowance List Test Program have been limited to submarines.) Submarine allowance lists have been approached by the Allowance List Test Program as having as their primary objective the goal of "maximizing" the vessel's effectiveness for carrying out a mission while being completely independent of all external logistics support. This correctly suggests that the aim has been to "maximize" the strategic value subject to logistical constraints. However, it is also true that the empirical research on submarine allowance lists has given attention to the "dual" consideration of "minimizing" logistical costs subject to strategic constraints. The several instances of quotation marks in the foregoing reflect the fact that these mathematical terms are used to suggest conveniently the direction of the empirical work; their employment can be precisely defended but this will not be required here. However, the problem formulations achieved by the Allowance List Test Program have progressed toward initial precision in another direction: that of improving existing operating practices in the Navy through derivation of quantitative policies capable of being applied to attain stated objectives. In order heuristically to show the above maximization of strategic value and its dual minimization of logistics costs as features of the Allowance List Test Program, one has only to consider application of results from the "military-worth" research.

Patterns of low and sporadic usage rates for the repair parts have already been described. After these empirical findings, attention was soon turned toward "military worth": how to determine relative measures of importance of having available one item rather than

another. In preliminary discussions with the Navy it was quite common to hear that nearly all installed components must have high military worth; otherwise they would not be on board. Thus, except for these few nonessential components, there would be little hope in trying to discriminate between military worth of repair parts. Neither of these guesses concerning uniformly high worth turned out to be valid.

An experiment was carried out under the Allowance List Test Program for one submarine, USS TIRU (SS 416) [30]. The basic assumptions characterized a sixty-day wartime patrol with complete isolation from all logistics support. For each one of the entire range of 1335 component applications, exactly one of four mission effects was selected as resulting from the component failure.

Code 1. The patrol would be terminated.

Code 2. The operational capability would be restricted.

Code 3. A moderate risk would be introduced.

Code 4. No restrictions on the mission would result.

It turned out that the almost perfectly consistent sets of answers supplied for the above "questionnaire" by three teams of experienced submarine officers yielded about 15 percent high-worth components (Codes 1 or 2) and 85 percent low-worth (Codes 3 or 4).

Next, about 32,000 wearable parts out of the 65,000 "pieces" in the submarine were selected as initial candidates for the allowance list. Experienced technicians evaluated the maintenance potential of each of the 32,000 parts by answering the following questions for each part.

Can ship's force replace the part?

Can the part be manufactured on board?

Are there standard jury-rig procedures for compensation?

Can the component operate without the part?

Can the part be removed intact without damage? (This provided entry to consideration of cannibalization from low-worth to high-worth applications.)

The simplest set of categories of importance for the repair parts was: A. no possibility of compensation for lack of part, B. compensation via manufacture or jury-rigging, C. no need for compensation, and D. ship's force cannot replace the part. Sets of military-worth measures then follow directly from combining mission effect and maintenance potential. For example, omitting "D":

High Worth: 1A, 1B, 2A, 2B

Low Worth: 3A, 3B, 4A, 4B, 1C, 2C, 3C, 4C.

Results for TIRU with the above scheme were that, excluding 5200 part applications representing equipage and "D," 7400 items (28 percent) had high worth and 19,000 (72 percent) had low worth. The original 65,000 pieces had thus been reduced to 26,000 allowance list candidates of which only 7400 were "high worth."

In summary, the highlights of the accomplishments to date are as follows (see also [29]);

1. Considerable insight has been gained into the problem of what sort of demand patterns are generated because of individual vessels.

2. The myth that "military worth" is a true imponderable was dispelled. It even was demonstrated that the research approach is feasible for wide application; in fact, the methodology now is in the process of being converted for operational application.

3. Usage data, population data representing the maximum range for usage, military-worth data, and item-cube data have been employed in a formulation of the allowance

list problem [2]. This formulation may be described as a method for employing these data to the goal of choosing, within a stipulated storage space constraint, that load of spare parts which best contributes to the accomplishment of the ship's operational objectives. It has been possible to formulate policies corresponding to specific objectives and then derive actual lists of items which best fulfill these objectives. For example, high military-worth items can be stocked independent of any past usage. Then all items can be considered for stocking on the basis of past usage alone, starting again with high-worth items and continuing until all items have been considered or until the storage space has been exhausted. The magnitudes of each item stocked in each of these "rounds" and the order in which the items are arranged into and within the military-worth classes are determined by the policies chosen for use in the formulation.

4. Several large-scale simulations have been performed for the Allowance List Test Program policies and methodology [32]. These have provided significant numerical results concerning the effectiveness of these policies, and in addition they have permitted an evaluation of the severity of the shipboard storage space constraint. The results regarding the space limitation (which indicate that space is not unduly critical) have far-reaching implications, not only for allowance lists, but for the distribution of stocks throughout the supply system. As for the effectiveness of the Allowance List Test Program methodology, it has been shown that the proposed methodology, if implemented, will result in a significant improvement over present procedures (i.e., minimize high-worth range shortages and significantly reduce shortages for all items). What is to be stressed is that these studies have not only gone forward along theoretical lines, but have also included extensive recognition of practical aspects, particularly in the provisions made for simulation. These have included formulation of data-processing problems resulting from changing populations, either for installed components or for stock-numbered items. Configurations change in time, as do the actual identification data for the installed parts. Also, there are related difficulties consequent to new equipment and parts which are entering the system — how does one transfer the methodology from the systems for which experience has been gained to the new and untried? These problems are but parts of the general concern which the Project has had for devising means to extract information from historical data.

5. Finally, plans are nearing completion for an operational test on TIRU, chiefly through numerical simulation of TIRU logistics vs various allowance lists and criteria therefor on the Logistics Computer, where it is planned to keep a record of future TIRU inventory histories.

LOGISTICS DATA PROCESSING

Extremely brief mention will be made here of the Project's accomplishments in the field of logistics data processing. Its role as a collaborator with the Logistics Branch of the Office of Naval Research as a pioneer in the field is quite well known. Several instances of application of the Logistics Computer have been cited above which, within the context of the problem areas, should give some of the flavor of the computational effort at the Project.

At its very beginning in 1949, the Project started analysis of requirements for a computing machine tailored to fit logistics. A representative set of problems were formulated in 1951 [24] which set specifications for the Logistics Computer which was received at the Project in February 1953. It is judged that the effort has been successful. First, although as in any computing establishment there has been wear and tear on programmers who have

applied their computer to problems beyond its originally planned scope, the Logistics Computer has been a vital tool for each instance of research described above and for many others. The Project was for many years (until 1958) exclusively concerned with research computing in which it was possible to get by with minimal input-output facilities and a relatively restricted capacity for machine instructions. This made it all the more necessary to try to get at the basic parts of a calculation and then make use of suitable ingenious factorizations. Then in 1958 a new contract was added which permitted more concerted efforts in analysis by numerical simulation. This broadened the scope of the computational efforts and made more demanding requirements on the computer. It has been possible to accelerate the computer modernization program in a manner responsive to these new requirements (see [25]). This has been most striking in the case of the FIIN Drum System, which has been described by Mr. Fennell [4].

The representative problems of 1951 which were used to fix the basic capabilities of the Logistics Computer have survived well in the light of subsequent experience. The main point to be noted in this connection is that principally the input-output requirements have turned out to be more demanding than originally estimated. Actually, with the original machine the input-output features were purposely limited in view of the primitive capabilities of the technology during 1951 - 1953. If one considers the effects of the extensive modifications and additions which have been made to the Logistics Computer and also some added details for the original representative problems, both the hardware and the underlying problems continue to be responsive to a large class of logistics data-processing tasks. These conclusions mainly follow from what has been accomplished on the computer problem concerning how to reference or address data in the memory (see [26]). This is the area treated by J. P. Fennell in [4].

BIBLIOGRAPHY

[1] M. Denicoff, Joseph P. Fennell, and H. Solomon, "Progress Report No. 1 — Requirements Determination," Logistics Research Project Serial 59/57 dated 4 June 1957.

[2] Marvin Denicoff and Henry Solomon, "Toward the formulation and solution of the allowance list problem," Logistics Research Project Serial T-84/58 dated 21 May 1958.

[3] Henry E. Eccles, Logistics in the National Defense, Harrisburg, 1959.

[4] Joseph Fennell, "An automatic addressing device," Naval Research Logistics Quarterly, Vol. 7, 1960.

[5] Joseph Fennell and Seiki Oshiro, "The dynamics of overhaul and replenishment systems for large equipments," Naval Research Logistics Quarterly, Vol. 3, 1956, pp. 19-44.

[6] Mina H. Gourary, "An optimum allowance list model," Naval Research Logistics Quarterly, Vol. 3, 1956, pp. 177-192.

[7] _____, "A simple rule for the consolidation of allowance lists," Naval Research Logistics Quarterly, Vol. 5, 1958, pp. 1-15.

[8] _____, "An examination of usage of ships parts for two destroyers," Logistics Research Project Serial T-89/58 dated June 1958.

[9] _____, "An examination of cruiser usage data," Part 1, Logistics Research Project Serial T-96/58 dated December 1958 and Part 2, Serial T-110/59 dated 1 September 1959.

[10] J. E. Hamilton, "Evaluation of data for the USS TIRU (SS 416)," Logistics Research Project Serial T-98/59 dated 5 January 1959 and Supplement dated 11 May 1959.

[11] _____, "Realism in empirical logistics research," Naval Research Logistics Quarterly, Vol. 7, 1960.

[12] _____, "Experience in data collection," a series of Logistics Research Project reports, in preparation, 1960.

[13] J. E. Hamilton and R. E. McShane, "The Korean Data Collection Program," Logistics Research Project Serial T-81/57 dated December 1957.

[14] James Hannan, "Approximation to Bayes risk in repeated play," Annals of Mathematics Study No. 39 (Princeton, 1957), pp. 97-139.

[15] I. Heller and C. B. Tompkins, "An extension of a theorem of Dantzig's," Annals of Mathematics Study No. 38, (Princeton, 1956), pp. 247-254.

[16] Ralph B. Hunt, "A mobile resupply concept to support a carrier task force," Logistics Research Project Serial 42/56 dated 9 November 1956.

[17] Ralph B. Hunt and E. F. Rosholdt, "The concepts of notional ship and notional value in logistic capability studies involving merchant ships," Naval Research Logistics Quarterly, Vol. 7, 1960, pp. 1-6.

[18] _____, "Determining merchant shipping requirements in integrated military planning," Naval Research Logistics Quarterly, Vol. 7, 1960.

[19] J. R. Isbell, "Finitary games," Annals of Mathematics Study No. 39, (Princeton, 1957), pp. 79-96.

[20] _____, "Absolute games," Annals of Mathematics Study No. 40, (Princeton, 1959), pp. 357-396.

[21] J. R. Isbell and W. H. Marlow, "Attrition games," Naval Research Logistics Quarterly, Vol. 3, 1956, pp. 71-94.

[22] _____, "Methods of mathematical tactics," The George Washington University Logistics Papers, Issue 14, 1956.

[23] J. E. Kelley, Jr., "A dynamic transportation model," Naval Research Logistics Quarterly, Vol. 2, 1955, pp. 175-180.

[24] Logistics Research Project Staff, "Some first considerations of requirements, features and questions relating to a proposed computing machine for logistics problems," The George Washington University Logistics Papers, Issue 4, Appendix II, 1951.

[25] Logistics Research Project Computing Laboratory Staff, Logistics Computer Manual, June 1960.

[26] W. H. Marlow, "Data processing for logistics," ONR Symposium Report ACR-28, 1958, pp. 47-58.

[27] S. B. Noble, "Some flow models of production constraints," Naval Research Logistics Quarterly, Vol. 7, 1960.

[28] Henry Solomon, "The use of interindustry economics for mobilization planning," The George Washington University Logistics Papers, Issue 15, Vol. II, 1957.

[29] _____, "The determination and use of military worth measurements for inventory systems," Naval Research Logistics Quarterly, Vol. 7, 1960.

[30] Henry Solomon, Joseph P. Fennell, and Marvin Denicoff, "A method for determining the military worth of spare parts," Logistics Research Project Serial T-82/58 dated April 1958.

[31] Henry Solomon, Raymond Lewis, and Marvin Denicoff, "Report on electronics and ordnance usage data for nine submarines," Logistics Research Project Serial T-97/59 dated 15 January 1959.

[32] Henry Solomon and Marvin Denicoff, "Simulations of alternative allowance list policies," Naval Research Logistics Quarterly, Vol. 7, 1960, pp. 137-149.

[33] C. B. Tompkins, "Numerical simulation of an operation," The George Washington University Logistics Papers, Issue 3, 1950.

[34] Nancy Whittemore, "The input-output technique as an aid in mobilization planning," The George Washington University Logistics Papers, Issue 15, vol. I, 1957.

* * *

RESEARCH FOR COMMAND LOGISTICS

E. K. Scofield

Long Beach Naval Shipyard

A review of the control responsibilities of a typical unified commander emphasizing logistic aspects. These responsibilities will be related to tasks, missions, employment of forces, and structure of the command organization. An indication will be made of what further progress is required, particularly wherein operations analysis and research methods can be applied. The suggested use of such means will include applications for improving normal staff and command functioning as well as hardware and facility development. The development and use of operations centers and direction systems will be included.

INTRODUCTION

In Dr. Marlow's preliminary outline for this conference, he suggested a discussion of logistic decision making. This is a full topic in itself. However, he went on to indicate that such a discussion also should cover plans for incorporating logistic matters in operational control centers. I am sure that Dr. Marlow had in mind the so-called control centers being developed for major commands such as CINCPAC and CINCLANT. At this level, it is most difficult to separate logistics and strategy. Even the operational control concepts, which these facilities are to accommodate, defy easy explanation.

Therefore, I would like to talk to you first about information and control—information being the means to an end, which is control. But, in keeping with the conference's theme of empirical methods in logistics research, I will make a few observations. Such remarks stem from experience in establishing the information and control centers in one unified command.

This effort was largely conducted in an atmosphere of pure empiricism—which is to say without the aid of science. I hope my discussion today will elicit some scientific research in certain aspects of that project. For this reason the title of my remarks is "Research for Command Logistics."

Accordingly, I will review the "state of the art" as regards command logistics with emphasis on logistic control. I hope to suggest areas wherein research and operational analysis methods may provide top command levels with means for information processing and control. In the sense used here, I mean control as the element which differentiates true strategic action from a haphazard series of improvisations.

I note that this conference marks the tenth anniversary of the George Washington University Logistics Research Project. For myself, it marks a completed cycle of some sort. Some years ago it was my privilege to expound on some of these same topics from this very platform. At that time, it was in the capacity of a Logistics Plans Officer, Staff, CINCLANT/CINCLANTFLT. Today, I speak without portfolio but with a recent background tour in the Pacific Command. Some of you may recognize a number of similarities in the situations reported then and today. We have carried forward considerable unfinished business.

This is not to imply that many people like yourselves, the continuing Logistics Research Project here at George Washington University, and others, have not greatly contributed to the

basic research. Your research has resulted in useful applications and technical developments in certain specialized fields such as inventory control and tactical direction systems. However, as of today, we have yet to provide either of the two top commands mentioned with command facilities adequate for the proper performance of command functions. It is true that some of the physical facilities are in the construction and installation stage. But I mean more than this. Allow me to explain.

For our discussions, I would expand the words, "command facilities" to include more than gadgets, elaborate "VIP" briefing rooms, and bomb-proof command centers. The often-used words, "know-how" indicate the important concept here that is lagging. The "know-how" I refer to is the technology of providing a commander-in-chief with the means for reviewing and evaluating information required to exercise strategic control of forces.

Our World War II experience, which could provide a point of departure, has not been perpetuated within the body and functioning of the successor headquarters. As a topic, information and control is largely undocumented. Now that wartime experienced staff officers are disappearing from the scene, there is little awareness that information and control is more than sending a man or a message. Today's cold war, occasional crisis, and possible hot war are situations which require far more comprehensive, rapid, and novel command actions compared with those of World War II. We need, then, a considerable amount of research and development in this business of providing the information and control tools for command. This task, however, is primarily one of applied research.

Much of the basic research has been done. The nature and structure of conflict are well documented. The interrelationship of strategy-tactics and logistics has been explored recently and with unusual perspicacity by Rear Admiral Henry Eccles. Further progress lies in applying available knowledge to the specific command requirements of our unified commanders.

Perhaps this area too long has been considered a sacred preserve of the military. I sincerely believe that you, as scientists and researchers, can assist in dispelling the "fog of war," to quote Admiral Spruance. If this can be done, we will go a long way towards putting strategic direction and control back in the responsible hands of our commanders-in-chief. In so doing, the expression, "unified command," will become a germane and accurate description.

While it may be just the fashion of the times, automated control centers for top commanders now are under construction. Such effort is belated. It is with far less specification of what they are to do compared with that which characterizes the control mechanism of numerous tactical-defense-type weapons systems. In the latter we have definite time and space inputs and outputs with data-processing machines in between.

Their task, of course, is to bring about the favorable employment of individual weapons or their vehicles either individually or, often, in a network situation. The Dew and Sage Line installations, the equipping of Navy ships with automatic data facilities, i.e., the Navy Tactical Data System, are familiar examples.

However, the task of designing control centers for a strategic direction system is not entirely a matter of communications, EDP, and automation. Other appurtenances include the tie in direction network, the communication linkage, and the command structure, real or fancied. These are facilities that must be designed as a whole, if we are to really assist the CINC and his staff to make and carry out logical decisions. Such decisions must be based on comprehensive and evaluated data, together with as much organized processing by the staff, lower echelons, or machines as is feasible. This, in turn, requires a high degree of compatibility in the

processing methods at various activities. Here I must add an observation. Such synchronization is not evident in today's staff work. We should realize, at this level, that information input, its processing and resultant output is often a variety of ideas, opinions, estimates, generalized missions, and tasks. Seldom can it be expressed easily in a numerical fashion.

I have long sought a descriptive word that connotes an organized information flow processing and system within and between the staffs of a command. It should represent a logic process that leads to decision making with its supporting actions. This is something more than standard doctrine or staff administrative instructions. At one time we had a little green book called "Fundamentals of Sound Military Decision." It was an analysis of the thought process involved in producing a logical military decision. Excellent in itself, it assumed that the staff and command was wired into the thought programming presented. I am quite certain this is not so insofar as it relates to adequately organizing the mental and logic processes of present commands. Some of this deficiency, insofar as facilities and hardware are concerned, may have been recognized by the present emphasis on developing command direction systems. What is still lacking, in my opinion, is an understanding of the cybernetics of a command direction system. I believe that such an insight is basic to the design of a direction system. Called for, then, is a comparative study of the human thought processes of the commander and his staff, which leads to the resultant products, i.e., plans, directives, etc. This must then be related to the ancilliary use of the more finite hardware items of the control system. With this analogy in mind, I would like to discuss decision making by top commanders and the role that information and control centers have in assisting command.

Note that several similar terms have been used. The first mentioned was an operations control center. Now this facility is referred to as an information and control center. I prefer the latter because, to paraphrase a truism, to him who is best informed, comes the ability to control. All the authority, responsibility, and organizational charts to the contrary, control is not synonymous with command. Control has to be acquired. Command can be exercised without control but not for long.

Let us be more specific about control. What is it? What is the framework within which it must function? What are the objectives of control? Finally, what are the tools of control? As we examine these aspects of control, I will point out where further research is needed to animate these information and control centers and their connecting direction system.

But first—what is control? Earlier, I quoted Dr. Rosinski in mentioning control. He writes more specifically by saying

> "It is the element of control which is the essence of strategy, control being the element which differentiates true strategic action from a haphazard series of improvisations."

I stress the words, "Haphazard series of improvisations." In the case of a unified commander, strategy and basic objectives have been laid down by national authorities. His forces have been allocated to him. His task is essentially one of improvising within his limited means and directed restraints. The sum of such improvisations, of course, hopefully brings about the desired and prescribed strategic action. Thus, the Commander-in-Chief must know what improvisations are possible and what is their individual and cumulative worth. Often as not the financial cost becomes a consideration. Furthermore, constant scrutiny by the public press and detail review by Washington authorities require elaborate feedback of information.

Since control implies strategic direction, the framework within which control is exercised is the entire command structure. Although my discussion today will be primarily concerned with headquarters' control centers, they are only intermediate hubs of a command direction system. Such a system is also under development.

Initial development of this system might be traced to the requirement for effecting atomic coordination. Originally it was recognized as a special situation. However, there is a growing conclusion that the structure represented by present command relationships is by itself inadequate for strategic control of other matters as well. However, I would hesitate before saying that information and control centers or direction systems are a substitute for traditional command and staff concepts.

Perhaps it is better to view this matter as being a technological modernization and improvement of the customary functioning of the command and staff. Note that I did not say command and staff organization. It, too, may be a candidate for change when integrated into a direction system.

It is fair to say that command requirements of modern conflict are not synonymous with the presently developed command structure. The ramifications of today's cold war with the always possible move into rapid and devastating tactical operations, makes former concepts of joint operations rather obsolete. At best these former concepts fit only a very particularized situation.

To summarize this introduction of strategic direction systems I would say this. Their purpose is to provide a framework within which the means, i.e., information and control, are available to exercise truly unified command from the President on down. A strategic direction system of this sort, if you will permit a gross analogy, is the wiring diagram to energize the recent Department of Defense Reorganization Act.

Having considered control as meaning improvisations of strategic value, I would like to simplify the definition by discussing "Objectives of Control."

OBJECTIVES OF CONTROL

Most simply stated, the first purpose of control is to provide for the comprehensive direction of the available means of force. Control also provides for the application of forces in a selective and concerted manner. It means allocation of the proper amount of effort to a variety of subordinate objectives which represent in toto the strategic responsibilities of the command. It might be said that a major purpose of control is to provide the means for informed exercise of command.

The classical objectives of control are listed in military textbooks somewhat as follows: Concentration of effort, Initiative, Decisiveness, Mobility, Direction, Surprise, Security, Confidence, Readiness, Economy of resources, and Harmonize military and political action. Some of these will be examined in a moment. However, to me, the above summarizes into this. The most important purpose of control is to translate assigned objectives into a series of nonhaphazard implementing and supporting actions which can be undertaken by the unified commander. The second most important purpose of control is to adjust these objectives and actions to the realities of a changing situation. This implies a feedback scheme in a very large measure. With a proper feedback of information, actions can be improvised that exhibit these listed objectives of control.

To further develop an understanding of what control implies, an examination of the "Tools of Control" will put the foregoing military philosophy into specifics.

TOOLS OF CONTROL

In a very true sense, control is a function of information. I would hasten to add, however, that raw information alone is not the answer. A simple tabulation shows the relationships between information and control:

Function	How
Information, collect	Staff, communications
Review	Staff, displays, briefings
Evaluation	Staff, EDPM assistance
Control	Staff, plans, SOPS, CINC
Decisions	CINC, staff for implementing direction system
Supervision, feedback	Communications, direction system, staff

This tabulation highlights the participation of the staff in the information and control process. Just how to organize the staff to perform these functions is a very big piece of the problem.

To go back to basics: The fundamental task of the command and assisting staff, particularly at unified command level, is the communication of ideas. The ability of these human brains to marshal ideas and convey them to others inside the staff and throughout the command is a real measure of command effectiveness. Each member of the staff adds, corrects, changes, adapts, and varies his work by interaction with other members of the staff, yet all work must relate to a unity of purpose. All the progress of the command, its control of the current situation, its plans for future contingencies are but a reflection of this ability to marshal facts and ideas with discrimination. Such a process is not in a vacuum but is guided by a flow of direction from higher authorities as well as information from lower echelons. Within the staff the hub for considering this information has become known as an Information and Control Center. When linked with similar centers throughout the command, we have what is called a strategic direction system.

In my opinion such facilities can produce a significant improvement in performing command and staff functions. This improvement is long overdue. It is, in fact, a modernization, including some automation of our command and staff functions. Without a lot of programming, however, these new facilities are as useless as their planned EDP computers. On the other hand, these facilities with their hardware are the physical tools around which effective control means can be instituted.

Since the staff is so involved in information and control, what are the facilities at a typical headquarters? There is, of course, the staff organization itself. The general pattern is to parcel out the work as follows (Figure 1):

By appending a few architectural symbols, this diagram would often be the space layout plan of the headquarters. I might add that the connecting corridors are still the major means of hooking up the 500 or so personnel of the staff, the so-called extra eyes, ears, and brains of the Commander-in-Chief. Out of recognition that this circuit has some serious impedance, the incoming-outgoing message traffic is duplicated in almost astronomical quantities and fanned out to everyone. There are the other tools of information and control, such as standard means to gear and synchronize the staff, i.e., conferences, trips, memos, etc. In total, all these coordinating measures leave much to be desired in providing for flow and evaluation of

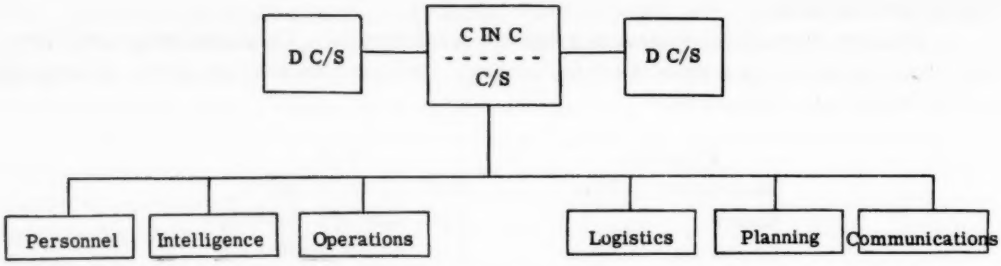


Figure 1 - Typical joint staff organization

essential information. I might add that there is plenty of extraneous information. Thus, the resulting staff effort is constantly directed to matters that have no real significance.

Occasionally some information on location of our own forces will be visually displayed in a plotting room. Over a period of time, general and specialized briefings for visiting VIP's will be given. If ever put together, this processed information could be considered the beginning of a stockpile of evaluated information for use of all the staff. However, this potential reservoir is too often held individually by members of the staff. It is compartmented to cover projects assigned to, or as is more often the case, projects initiated by individuals. This coverage is largely unprogrammed in the sense that a definite scheme and procedure is applied over a period of time in the direction of recognized courses of action.

Despite this rather bleak picture, staff functioning is susceptible to improvement. I refer to the use of a visual-display type of information center. It can be a real integrating force in the staff effort.

I stress the words, visual display. This, in itself, forces summarization. More important, information displays of this nature have the great advantage of providing pinpointed group discussion leading to review and evaluation. Individuals in each group represent a specialized interest of their parent staff division.

At the moment, I am talking more about static type information displays rather than electronic displays of moving situations. In other words, permit this simile. These displays can largely represent a stockpile of basic evaluated information relative to the command. Here is the reference data which is the basis for planning and then revising plans to meet specific contingencies as they unfold. In fact, the planning process itself can and should be the major source of stockpiled information.

Inclusion of the planning displays in an information and control center is most important. Strategic control and so-called operational control in a dynamic situation, in my opinion, is exercised through planning and through the control and adjustment of the ensuing action.

I consider this topic of planning to warrant a thoroughgoing research and analysis. Such a review should cover not only the techniques in summarizing and displaying plans in an information and control center, but should also include an examination of the entire planning process within the command. In this context, planning is a basic feature of the over-all strategic direction system.

Peacetime planning for selected likely contingencies is, of course, reduced to written plans throughout the command. More often than not, I fear they are an academic exercise for a few unfortunate individuals. Generalities, compromises, and a plethora of assumptions, usually unrecorded, are typical of peacetime plans. There seems to be an aversion to developing supporting detail from whence feasibility can be assessed. Even if detail in supporting plans is developed, the process of aggregating it for feasibility consideration is quite beyond the available staff talent and capabilities.

I suppose there is in the back of every planner's mind the depressing thought that the monumental work required for a complete plan is largely wasted effort. Contingencies as planned, seldom, if ever, materialize. So often the applicable plan is not even adaptable as a point of departure.

What is the remedy for this situation? I suggest that planning in peacetime has a real mission. It should stockpile information. It should provide for a continuing information flow which will permit stockpiled information to be melded with the current situation. In military terminology, the end result should be a capability to convert operation plans at any time into operation orders.

A few more words about planning. Peacetime planning can be considered primarily a regenerative cycle. Part of the difficulty is the lack of feedback and then the difficulty is what to do with it. With present planning cycles taking up to four years, there can be obviously little continuity in such a cycle. During that time many shifts of planners with their learned insight have left the scene—to say nothing of changes in the Commanders' echelons.

Theoretically, a horizontal regenerative process on the planning effort should take place within the staff. Here, modifications in one part of the plan should induce changes in other parts. In a vertical sense, supporting plans, together with their review are also part of the regenerative process. Supposedly, deficiencies are ascertained. Improvisations of some sorts are determined in order to present a course of action that is feasible. It might not be suitable, or even adequate, but at least one would know that it is possible.

The foregoing description of the planning cycle may be good theory. But it is working on such a small scale, that I question its validity, both as a theory and its present usefulness in providing a means of control. It is for these reasons that I consider a priority research task to be an examination and improvement of our present planning process.

We also need to find a way for a reasonably rapid production of plans including concurrent development at various organizational levels. There is required a means of daily review of progress. Once developed, a means of maintaining plans ready for the designed contingency is a further requirement.

Contingency plans can be a living dynamic expression of the command's mission: locked up in a series of safes or inside a few planner's minds makes them academic museum pieces.

From these observations I consider it most necessary that any control and information center should provide space and procedures for the development, review, and updating of plans. The data preferably should be in summarized form suitable for visual presentation and correlation.

If I seem to emphasize visualization for group participation, it is because we are dealing with human computers. The result is supposed to be a product of logic. That the process of programming these human computers is not always compatible is another reason for including a portion of the planning function in the information and control center.

There is no doubt that the planning function is an essential tool of control. When the nature of problems is thought out in advance, there is likelihood that the unanticipated features

of the contingency will be handled by wise improvisation. A ready stockpile of applicable planning information can be developed ahead of time, whereas the exact contingency never can be fully anticipated.

Experience has shown that without this preparation, any planning after the "flap" starts is never fully effective. It is always wasteful of resources. We have been lucky in that recent international crises have been one at a time. What we should try to do is to establish control at the beginning of the operation rather than midway.

Perhaps the foregoing has largely been a review of a particular planning philosophy. Indicated in any event has been the relationship of control to strategic direction. Control in large measure is a function of dynamic plans based on a stockpile of updated planning information. I have outlined the role that an information and control center facility can play in integrating the staff work. Visual displays have been stressed as a means of obtaining individual staff participation. Indicated, too, has been the need for continued research towards improving the entire planning cycle.

Before expanding our discussion on research, a schematic view of an information and control center may be helpful. While we are not architects, this diagram of the physical aspects of an information and control center is useful. It is the stage where the output of our information process is displayed and heard. It also provides means for audience reaction to affect the stage show (See Figure 2).

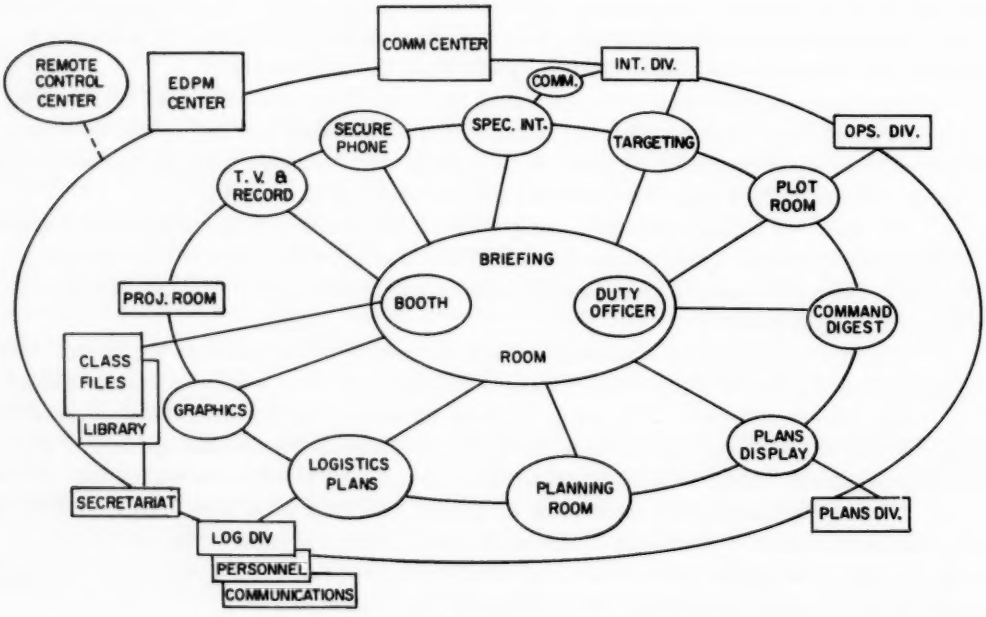


Figure 2 - Typical information and control center

FACILITIES FOR CONTROL

Strategic Direction System

The information and control center just described might be envisioned as a hub, or central exchange of the command's strategic direction system. Such an area system is in turn considered to be part of the nation's world-wide military strategic direction system.

I often find official definitions difficult, but allow me to quote one which is descriptive.[1]

> "Strategic direction systems are those command and combat direction systems employed by the unified commander, the subordinate unified commanders and the service chiefs whose headquarters are ashore. These systems comprise the JCS War Room, the Joint War Room Annex, the Chief of Staff of the Air Forces Command Center, The Chief of Staff of the Army's Command Post, CNO's Flag Plot, and the Command Centers of the unified commanders, together with their internal information and data processing equipment and the communications and data links which interconnect and tie them to various tactical combat direction systems."

If we had such a system, I could conclude my discussion. But the fact is that there is no completed system. Only recently has it been recognized that more than compatibly linked hardware is required. Even the latter is difficult, requiring study and development of a common ideographic language, standardized plotting and display techniques, and a uniform combat reporting system. What is more difficult and most important is formulating operational command concepts for a comprehensive and responsive system through which intelligent control of military power may be exercised.

Various scientific organizations are now investigating both aspects just mentioned. Some of these tasks will never be finished so long as technological development continues in weapons and other means of attaining strategic objectives. The point here is that a variety of operational control concepts must be envisioned to accommodate a number of different tactical situations. Here, once again, as Admiral Eccles would point out, is a blending of tactics and strategy.

There are other considerations which should be kept in mind when formulating concepts for control systems. We have talked about certain tools, such as planning. Now we have introduced facilities, such as information and control centers and direction systems. Further on I will elaborate on some case examples of certain specific problems where control depends upon information processing. However, before examining those problems, a few remarks should be made concerning the atmosphere in which the formulation of concept for a strategic direction system finds itself.

At the outset I wish to make an observation. It is one which has been relabeled occasionally as an unwarranted assumption. The very fact that we are considering superimposing a direction system on a command system is not an anomaly. It is recognition that the formalized command structure is an inadequate vehicle to provide comprehensive and responsive control of military power. A comparison of the normal command structure and functional task assignments will indicate why the former is not always conducive to exercising direction or control.

Keeping in mind our rough analogy that control lines might be envisioned as a wiring diagram for the exercise of command, the question arises, can we describe what has to be hooked up. Here is one hook-up which is the normal way to describe a unified command.

The Command Structure

This structure (See Figure 3) provides for command channels quite useful for administration, logistics, training, and general housekeeping of the military forces in the command. One might epitomize this structure as the garrison organization of the command. These functions, incidentally, largely by-pass the unified commander and follow the independent dollar paths associated with each service.

What is the capability of this organizational structure to function as a strategic and combat directional system? It has some capabilities as was proved in World War II. Direct control lines are available to forward area garrison forces (subordinate unified commands) or composite forces establishing a new garrison area, (joint task force (JTF's)).

Through his component commanders, the unified commander can control task assignments of a single service nature. Insofar as such tasks need no correlation with other services and forces, a directional system easily follows normal command lines.

Unfortunately, JTF's and subordinate area commands can no longer be envisioned as the vehicle for directing the major striking power of the theater.

Major striking forces today are unlimited-war forces, with atomic capabilities. They, of course, may support and back up limited-war situations characterized by defined geographical limits. It is rather inconceivable that these striking forces would be placed under the operational control of local area commanders, when the forces' mobility and range encompass the entire theater.

Except for the ASW campaign this dimension was not with us in World War II. Today we have a situation calling for a variety of command arrangements. They can be just as difficult of predetermining as attempting to preplan for all possible contingencies.

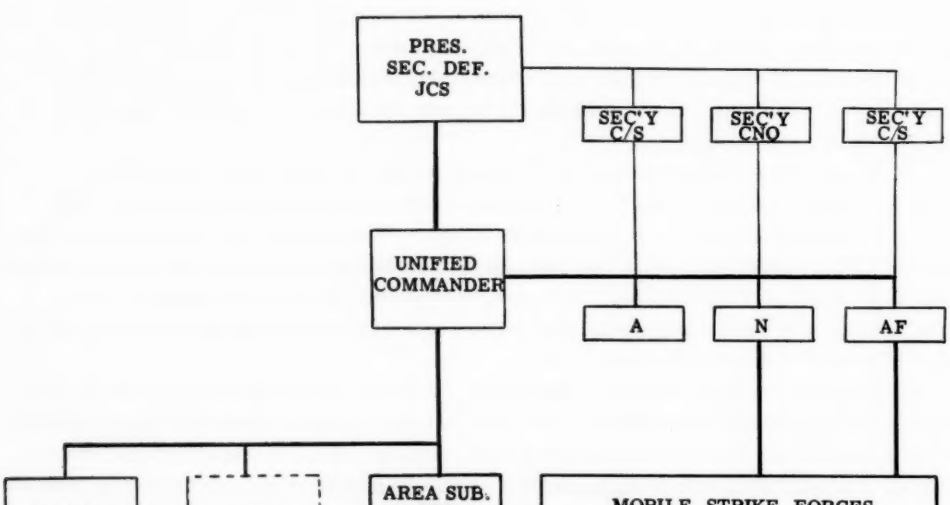


Figure 3 - Typical unified command structure, 1959

The major operational problem is the coordination of the mobile strike forces when their operations are in mutual support of each other. This is particularly true when different services are involved, operating from widely separated bases or headquarters, yet perhaps striking the same target areas. To insure that such strikes will be in concert, operational control of some description inevitably gravitates to the echelon of command where the control is vested in a common superior.

This same concerted action would seem to be also a requirement under other than full-scale war conditions. Back-up support to area commands such as subordinate unified Commanders or JTF's may well require directed coordination by their common superior, i.e., the CINC in limited types of operations.

In these situations the command requiring the support is directly responsible to the unified commander, but his back-up assistance is responsible to different intermediate commanders who in turn report to the unified commander. Difficulties and conflicts inevitably arise and naturally some filter up to the common superior. He then finds himself in a position of having to arbitrate in detail on many tactical coordination measures (See Figure 3).

Even worse, since his field commander has no logistic resources of his own, he must depend on the parent services of the forces allocated to him. Past experience has shown an inevitable flood of detail logistic considerations coming up directly from the subordinate unified commander. Often as not these requirements have to be reconciled by the CINC, which means intervention with his component commanders. To say that the unified commander is only concerned with broad policy and strategic direction is not in accord with the facts of life.

The cumbersomeness and inadequacy of using normal administrative command channels for operation control is certainly obvious to those of us who have sat through recent "flaps." Each service is fairly well set up to run its own war with its own forces. This is because information flow, command lines, and operational control lines are parallel, if not one and the same, insofar as one service is considered.

When these lines are spliced into the unified commander in order to provide for coordinated employment of striking forces with those allocated to subordinate unified commands, the result isn't always that which is desired. The spliced lines often carry incompatible voltages. Which is to say the information flow is inadequate, unprocessable, and too often insignificant. Naturally decisions made on this basis and passed down through these same high-impedance splices lose much of their effectiveness.

In partial recognition of this situation and with the hope of decentralizing coordination action to the forward area, local coordination centers have been considered. The thought here is that representatives of the striking forces would obtain and exchange information and relay recommended coordination measures back to their parent strike-force commanders. No command or directive authority would be vested in these information exchange centers. Recommended measures would be entirely on a voluntary-acceptance basis by the separate strike commanders. If and when any disagreements became apparent, referral was supposed to be to the unified commander for his resolution with component commanders. I consider this a tenuous, ineffective control device, if for no other reason than it is control by voluntary cooperation.

One way of summarizing the current operational-control system in what I consider a typical example, is to take a look at the unilateral planned development of service direction systems. Obviously, if we could properly hook them together, the unified commander could obtain a strategic direction system adequate for performance of his command responsibilities.

However, this hooking up is like adding oranges and apples. This is illustrated by looking at the hub of the various component commanders' direction systems (See Figure 4). Development to date has been strictly along unilateral lines, at least in one unified command. Figure 4 typifies this unilateral service approach. To me it also illustrates the antithesis of unity of command. A common 800-line telephone exchange is the major concession to integrated information processing. Separate and individual EDPM installations are envisioned, perhaps compatible, but not by design. Unilateral programming, duplication of intelligence, plotting, and targetry functions are planned. The fact that all of these activities are under one roof has not brought about any particular integration except for housekeeping. But since these centers are under one roof there is a myriad of possibilities for developing an information and control system adequate for the unified commander to perform his command functions either directly or through his component commanders.

Here is what could be done and what should be the goal. In Figure 5 one sees how the several service direction systems might be integrated by a common operations control center. It provides for a variety of command arrangements. There are other variations. However, if we accept the trend of centralized control as being a valid requirement, then what is pictured

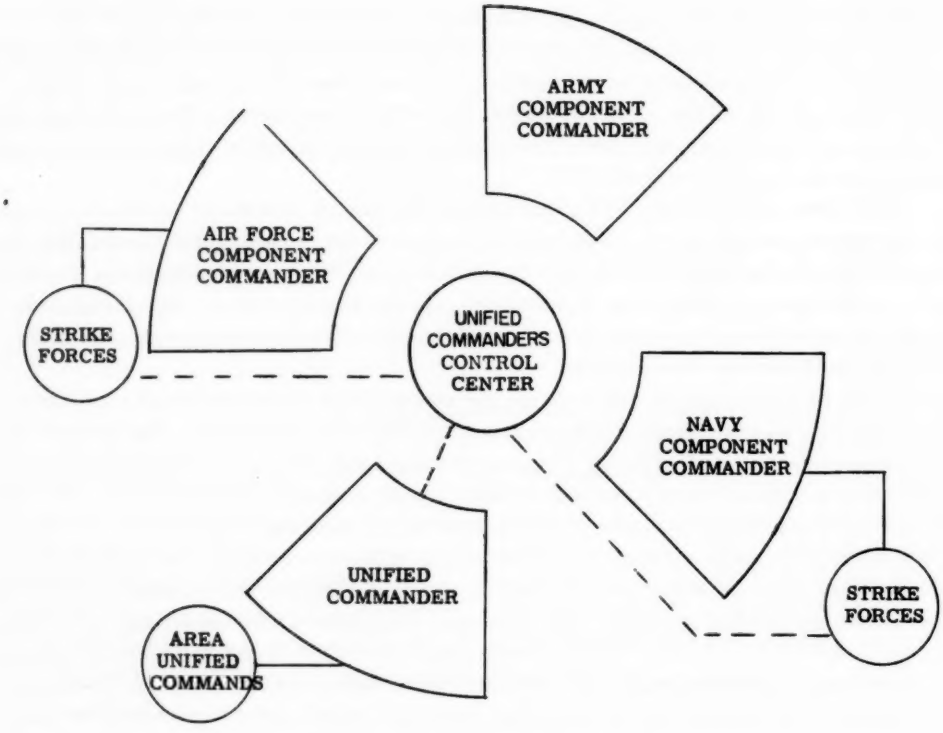


Figure 4 - Service concepts of unified command direction system

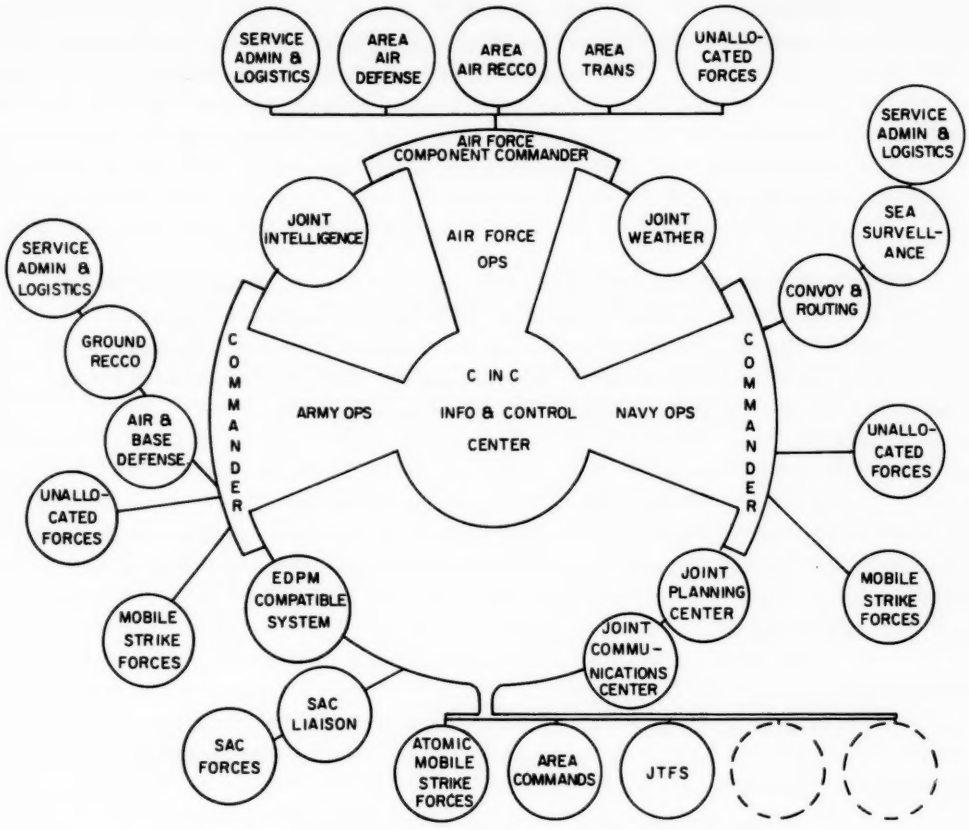


Figure 5 - Proposed OP control system

here provides the means for component commanders (whether they want to or not) to contribute information and be in the channel of control (which they do want). In short, it provides for an integrated information and control center as follows:

Function	How
Information, collect	Common-language reporting; cross-connected communication systems.
Review	Standardized display formats and symbols.
Evaluation	Integrated and compatible information processing.

Function	How
Control	Command planning system; personal presence of component commanders for advice and acting as CINC deputies, as well as commanders in own right for specialized operations
Supervision	
Feedback	

When hooked up with the tactical data systems of operating forces and subordinate commands, it may result in the millennium of producing a truly unified command.

I believe that such a hub would be an adequate fulcrum upon which the unified commanders' strategic combat direction system could be built. Such a direction system would meet these general requirements which are considered basic:

It must be:

1. ADEQUATE
   - A. For performance of the CINC's command functions.
   - B. To handle large volumes of information; to enhance reporting, consolidating, summarizing, review and display of this information, machine data-processing should be envisioned.
2. COMPATIBLE
   - A. With command facilities (i. e., information and control systems) of components.
   - B. With JCS, adjacent U. S. Commands and control system of allies.
   - C. With main control centers remoted from Administrative Headquarters as normal mode of operation.
   - D. With capabilities of present or planned communications systems.
3. FLEXIBLE
   - A. To provide directional capabilities under various conditions of international tension for peacetime, limited war, or all-out war.
   - B. To provide directional capabilities for atomic or nonatomic situations.
   - C. To provide alternate operational control sites for CINC or Successor in Command.
   - D. To accommodate a variety of command arrangements.
4. INTEGRATED
   - A. With the theater's regular command activities so that it can immediately function in event of increased emergencies.
   - B. With present or planned communications systems.
5. CAPABLE OF GROWTH
   - A. Recognizing the increased operational command responsibilities of unified commanders as result of DOD Reorganization Act.
   - B. To provide directional capabilities for employment of future weapons systems.
   - C. To provide minimum disruption of present capabilities during establishment period.

My remarks to this juncture have all been introductory to the main topic and title of the presentation. We have reviewed the control responsibilities of the unified commander with some indication of what control is. Objectives, tools, facilities, and the organizational situation have been related to control requirements. With this background, I feel now we can appropriately discuss some research projects for making logistic control measures effective. Of course, at this level, command logistics merges with other command functions and it is fruitless semantics to attempt precise partition of one from the others. Therefore the first research project I will cover is Readiness-Posture.

READINESS-POSTURE

The mobile nature of the major striking forces in a unified command, particularly fleet units, is an advantage requiring constant surveillance. These forces are a very significant portion of our retaliatory capability. Their optimum deployment pattern for an all-out war situation seldom coincides with favorable deployments for the lesser but more likely contingency operations. We would be foolish to maintain continuously any optimum positioning of Fleets and Air Forces since this would be exchanging mobility for a "sitting duck" posture. In peacetime, men (and their ships) return to port, planes rotate to advance areas and return to home base. Thus, for a majority of time, a compromise is accepted as opposed to maintaining a maximum readiness position for general war. The value of this compromise is further affected by variable weather conditions, changes in force strength, and the efficiency of units. The latter is mentioned because each unit deployment to forward area is initially in a final training and shakedown status.

Even if all these factors cannot be absolutely measured, their resultant from day to day can be a comparative expression of the command's emergency posture. How to state this condition relatively is the question.

I feel that the first item to be covered in the daily briefing of the Commander in Chief should be an analysis of today's capability compared with yesterday's. This should be in meaningful terms based on a standard representing the maximum attainable capability. This latter is the emergency war plan itself. Such a plan by definition is supposed to be a capabilities plan.

Naturally, it can be assumed such a plan includes allocated weapons, designated forces and geographical take-off points. Thus, the problem of analyzing and comparing relative postures is primarily one of processing available data. To accomplish this, some basic research towards developing a rapid and scientific means of computing and comparing readiness changes is needed.

The results should be presented in visual and summary form. Further, the presented capability should be related to courses of action that could be taken to improve the situation. Such information, properly assessed, could be the basis of controls of real strategic worth to the CINC.

For example, I believe such an analytical situation report could be a rational and precise basis for requesting augmentation of forces to the theater or temporary assignment to SAC of theater targets. At the least, such an analysis and accepted course of action would become a calculated factor of the so-called calculated risk.

In discussing this particular matter, the subject of compatibility arises. The JCS have the same problem on a world-wide scale. To arrive at any such determination of readiness means that the terms used by each unified commander to express his theater's readiness are

terms which can be aggregated in Washington. This means, of course, standardization and a common reporting language.

It is the old story otherwise, of trying to add apples and oranges. This has happened. The resultant readiness picture was off course, erroneous and seriously so. Perhaps a greater concern would have been evident if our retaliatory force wasn't plentiful, compared with minimum needed. However, such plenty could be reduced by fortunes of war. Then our ability to undertake retaliatory measures, after having taken the first blow ourselves, would seem to call for precise direction of remaining strike power. This, of course, implies continued development of an accurate readiness reporting, evaluation, and presentation system at top command levels.

Our readiness posture is more than an assessment of our own capabilities vis a vis the enemy. For a number of years this country has been cultivating allies and friends. Their readiness and capabilities can contribute to our strategy. This, of course, has been recognized. I next would like to discuss this matter, as an example of wherein a considerable amount of secondary research could be helpful in utilizing this potential.

INDIGENOUS FORCES–CAPABILITIES

One of the tenets of our long-term national strategy has been the equipping of smaller nations in order to strengthen their self defense. The Military Aid Program has focused on achieving internal security plus sufficient holding power to temporarily stave off external military aggression. The holding capability should be sufficient until relief could be furnished by mobile U.S. striking forces.

Aggression or threat of it may not warrant full-scale involvement of U.S. Forces, if some of the border incidents of the past are recalled. Lesser measures of special logistic support, such as on-the-spot training and advice can be keys to realizing the potential capabilities of our ally to effectively resist in specific incidents.

The Military Aid Program is in large sense a Logistic Program. It is unique in that the unified commander is the manager in certain details of this effort, rather than the service component commanders who normally handle this type of detail for their own service. The unified commander has the task of reviewing, sometimes determining detailed requirements, country by country, service by service. He justifies them to DOD and even before Congress. Implementation of the hardware portion of the program, i.e., the procurement and distribution of goods to aid countries, is done by the logistic systems of the three U.S. Services. These pipe lines are departmental and often quite aside from those supporting U.S. Forces in the theater. The latter pipelines are naturally responsive to component commanders of the unified command. This explanation of the Military Aid Program is recited to indicate that the unified commander has no easy direct means within his own command structure to make emergency revisions to programs. Such action, including the funding, procurement, and rapid delivery of specific reenforcing items is often the key to effective resistance by an ally under active military pressure. In these instances the normal 24 months or so lead-time characterizing military aid deliveries is out of the question. Another way of viewing this situation is to consider the similarity to our own contingency plans calling for direct involvement of U.S. Forces. Just as such plans are unlikely to dovetail exactly or even recognizably into a real situation, so is our M.A.P. equipping and training program unlikely to be at a point where it can always cope adequately with the "flap" in hand. Specific key items, by that I mean

equipment, training, and technological deficiencies, will exist. These, experience has shown, have limited or even prevented the particular M. A. P. country under fire from taking immediate positive resisting measures which may have one potential capability.

In recalling one crisis, some of these key deficiencies came to light after the situation became active and serious. Ignoring for the moment certain politics that are always involved in Military Aid Programs and intensified when emergency requirements come rolling in, I would cite a few examples of bona fide key requirements. The opposing strategy of this particular crisis was the interdiction of the so-called offshore islands. Ours was an opposite one—prevent the isolation, blockade, and perhaps abandonment. Supply runs to certain islands, either by sea or by air drop, became the scoring play of the day.

Although the indigenous forces had some navy amphibious ships, sufficient landing craft were not in their possession. In the case of air resupply they had plenty of cargo aircraft. In both cases they allegedly had not the technical skill to accomplish over-the-beach or air drop deliveries.

The point I wish to make here is that these deficiencies were not entirely unknown but they had not been translated into planned improvisations for an operation that had been long and widely anticipated. A rather frantic search of U.S. resources in the area ensued before reserve landing craft were located and forwarded. In fact, we peremptorily off-loaded boats from some of our own amphibious ships.

In the air, too, a rather hectic time was experienced in tailoring and moving forward an aerial resupply company (U.S.) to train and advise the local forces. Cargo parachutes and associated equipment became a prime concern.

To make matters more difficult, every move was contested by comptrollers where service funds or equipment was involved. If the matter was Military Aid, only the Department of Defense could approve diversions or revisions. Finally, matters became so critical that these monetary and centralized approving controls were relaxed. There followed, of course, a snow-ball effect, with each U.S. Service putting its best propaganda foot forward by trying to get in on the show (incidentally charging everything to Military Aid). The early termination of the crisis and service commitments elsewhere prevented this volunteer bandwagon from getting entirely out of hand.

In my opinion, the information and control process available to the CINC was not adequate. Prior planning in this most anticipated contingency was hardly worth even reference value. Situation reports, particularly logistic ones, did little to clarify progress or substantiate requirements. Requirements, themselves, originating in the field were not based on a common understanding of concepts and available forces.

This account may be harsh and critical. "We made out didn't we?" is an easy rejoinder. However, I am one who believes that uncontrolled mass improvisations of which only a few provide the substance for the real strategy may not always be available to us. Suppose this had been a contingency at several locations, or even a minor limited war involving U.S. Forces? Could we have judiciously utilized our resources so that each counted to the fullest?

I honestly believe that certain stockpiled information at the CINC's Headquarters would be invaluable. For example, we have the means in our Military Assistance Advisory Groups to assess the state of training and progress of equipping as it applies to each of our Military Aid Programs. These deficiencies exist now but are to be remedied in plans for future programs. Therefore, these current deficiencies can be summarized and translated into major

items, such as supplementing U.S. Training Teams, etc., which would be immediately required to make effective partially-trained and equipped resources.

On the other hand, the CINC should have a better catalog of such U.S. resources within his own command. For example, something similar to our advanced-base catalog of functional components, could be developed for the skills and equipment of functional reenforcing teams. Capability of tailoring such elements should be considered in order that varying degrees of indigenous support will not be duplicated. We must be prepared to commit such equipment and specialists on a piecemeal and immediate basis. Only by so doing will our scarce airlift and sealift be utilized for those items which literally are sometimes "the missing nails and horseshoes, for lack of which the battle is lost." So much for an inventory cataloging problem. To conclude my comments on possible research programs, I will discuss Atomic Coordination.

ATOMIC STRIKE COORDINATION

For obvious reasons, this topic can be discussed herein only in generalities. It is safe to say that over the years a very considerable amount of preplanning has naturally taken place. As atomic weapons became more plentiful, their sole employment for strategic purposes by a single delivery system has expanded to cover lesser strategic and even tactical objectives. Delivery is now possible, of course, by many means under numerous commands.

This situation, even if it represents an "overkill" capability, is no reason to forget the original control objective of precise utilization and maximization of effect. In the event we are the recipient of the first attack, it could be a vital necessity that we be able to maximize our remaining atomic capability.

The present laborious preplanning is a large volume data process. The product is one typed to a specific situation. Electronic-data-machine-processing methods properly programmed and compatibly so between various headquarters show promise of better handling of the basic planning task, together with rapid revision. Much applied research can be directed to this area, including that of rapid communication links, in order to improve the coordination tools of the Commander in Chief.

However, the greatest of all difficulties in this task of coordination is the matter of command relationships. Closely allied is the topic of operational control. Tactical, supporting, strategic limited, general emergency, functional assignments, are all situations sufficiently different to warrant, I believe, tailored and special control methods.

One conclusion is most plain to me. Whether control is directly by CINC or by task assignments to component commanders, the scheme must be proclaimed by a clear-cut voicing of the CINC. Unless so enunciated, that is, so long as component commanders are not told clearly that the CINC will directly control strike forces, these commanders are compelled to retain direction systems and information and control facilities permitting them to exercise control in their own, and probably unconcerted, ways.

The possibility of unilateral, uncorrelated development of individual service direction systems could result in an incompatibility, delaying if not making ineffectual, some of the basic tools for exercising unified command. We are in an age where control measures, i.e., exercise of command, coordination, or management of military operations, is as far removed from directing a World War II Joint Operation as a bazooka is distant from an ICBM.

In conclusion, it is with these thoughts in mind that I suggest to you scientists and operations analysts, a refocusing of your sights. Too long have we put our limited talent and effort on lesser matters which are more susceptible to numerical description. Top command

information and control problems warrant the same expert assistance. I am certain that command problems and command logistics are also susceptible to the same analysis and subsequent improvement to which you researchers have contributed so much in supporting fields of interest.

* * *

SOME ASPECTS OF MATHEMATICS IN SOCIAL SCIENCES

C. Tompkins

University of California, Los Angeles, and
Institute for Defense Analyses, Princeton;
Consultant to the Logistics Research Project

The ultimate aim of this paper is to examine the compromises which can be made in order to reach a point where measurement and computation are feasible in some of the quantitative aspects of the social sciences.

The basic mathematical theories upon which the social sciences depend include:

1. Linear attrition theory including Feller's abstract formulation of parabolic equations.
2. Utility theory as a branch of the calculus of variations, described, for example, by G. E. Evans.
3. Combination of utility theories and competition in many play games, including Bellman's theory of dynamic programming.

Basic empirical aspects of the social sciences include:

1. The acquisition of information about probability distributions, such as distribution of demand.
2. The testing of hypotheses related to reality, such as a hypothesis of rational behavior as formulated and tested by J. Marschak.
3. The intelligent formation of a priori Bayes probabilities and their correction as demanded by L. J. Savage in his The Foundations of Statistics and as carried out by W. A. Smith.
4. Observations of value of various competitive situations, based on simulation, experience, or other valid means of practical realization of quantities introduced theoretically in dynamic programming studies.

Practical considerations require action after limited observation of existing conditions, and limitations of computational facilities may prohibit complete numerical analysis of situations whose underlying parameters are believed to have been accurately evaluated. This leads to an almost epistemological question as to what can be learned through aggregation of data related to seemingly similar situations and to difficult mathematical questions of reducing the computational extent of problems. Some means of examining these questions will be treated, but the paper will not attempt to present a unified theory covering any substantial set of these problems.

INTRODUCTION

My purpose here is to list some of the fields of mathematical development which seem to be of urgent importance to realistic quantitative studies of social sciences. In this I shall try to emphasize the realism of the study, although it will be obvious that I must allow this realism to be more of a constraint at some time than at others.

More explicitly I shall view the development of any science as a cyclic process of observation, abstraction, analysis, and application. The observations are made of reality

and the applications are made to reality. The abstractions are made from reality, and the purpose of the analysis is to lead from abstract assumptions based on observations to results which are simply applicable to reality. The purpose of the application may be utilitarian in the case of an applied science or a test for anomalies or contradictions in the case of a developing science.

Much work has been done during the past few years on the development of tools for analyses which are useful in this program. Thus, linear programming, for example, is a powerful tool for analysis; however, it becomes a realistic tool only after coefficients have been properly enumerated and estimated and these processes checked by adequate experiment or observation.

I shall describe some analytical tools below, but only as examples. These include linear attrition theory leading to sets of linear differential equations, part of utility theory described as a branch of the calculus of variations, and a development of this utility theory into the modern theory of dynamic programming.

Proper application of these and other similar tools to realistic situations demands compromise between two conflicting demands—a demand for sufficiently extensive sets of dimensions and parameters to guarantee a reasonably faithful representation of the reality depicted (on the one hand), and a demand for restricted numbers of dimensions and parameters to permit reliable measurement and reasonably economical computation (on the other hand). I shall turn my attention to this area—to methods of attack which may be expected to help resolve this competition of demands. My remarks are more properly to be interpreted as an indication that there remains much valuable work which can be carried out by groups like the Logistics Research Project than as contributions with any scientific significance of their own.

Incidentally, I should like to note that I was greatly influenced by Professor J. Marschak, and would suggest that readers consult his recent works for lucid and precise treatments of various topics using techniques similar to some outlined here.

SOME EXAMPLES OF ANALYSES TO BE SUPPORTED

A principal symptom of the development of precise analytical treatment of abstractions of social systems has been the requirement for a more precise statement of the nature of the imponderables of classical theory. To this end, in the modern versions of old theories, it has been instructive to introduce probabilistic distributions from which random variables are drawn. It has also been necessary to describe more precisely what is to be accomplished and what will be accomplished under various patterns of action. In these descriptions the probabilistic nature of social reactions is taken into account where this seems necessary, but this probabilistic nature is to be described in terms of indicated parameters and not shunted off with a label of imponderable.

In all this, the difference in meaning between the words precise and accurate (as used by numerical analysts and applied mathematicians) should be borne in mind. A precise statement is not necessarily an accurate one in the sense that it gives a faithful picture of reality; rather it is a well defined statement from a quantitative point of view. In classical literature the value of the ratio between the circumference of a circle and its diameter (the constant pi) was stated with a precision of 720 places to the right of the decimal point. As a matter of fact, there was a mistake in the 515th decimal and the accuracy of the statement was a couple of hundred of decimals less than its precision. In much of numerical analysis extra precision is

carried to protect what accuracy is present (frequently ten decimals are carried through approximate calculations in order to assure seven correct decimals at the end) and in much of science precise statements of models are made and studied whether these models accurately depict reality or not. In particular, the models are described in many cases through parameters which are represented by literal marks and analyzed algebraically, and part of the application of the model (if it can be applied) is the determination of numerical values for these parameters. In some cases the theory can be rejected on the ground that there is no possible way to determine numerical values of the parameters which would give results agreeing in any way with experience; in other cases it is almost obvious that there must be some values which will work well enough to agree reasonably with all reported observations.

The realistic application of theoretical models requires the determination of parameters and the testing of the hypotheses of the theory. In this section I shall outline a few more or less typical theories which have been presented as analytical tools. My purpose is to have these theories available as examples in later sections.

I shall consider a particular method of attack on competitive situations in which there are two competitors. The classical consideration of strategy, tactics, and intelligence will not arise explicitly in the analysis here, but ideas with similar roles and subject to more precise theoretical definitions will dominate the analysis. In particular, it is convenient to distinguish between three factors in the quantitative studies of some realistic competitions. These factors are:

1. Computation of the probabilities of all plausible results of actions which might be undertaken by the competitors.
2. Computation of the utility of any configuration or continuing pattern of configurations in time.
3. Determination of an action or a course of actions which should maximize one competitor's expected utility among all actions possible.

I shall sketch the development of an example of each of these factors below and later indicate some wider applications.

A general technique for providing a probabilistic model for the results of actions, factor 1 above, is available in some recent work of W. Feller [1], [2], [3], and [4]. This attack requires that the possible configurations which may be generated be arranged in some (not necessarily one-dimensional) order so that passage from one configuration to another is plausible in a short time only if the configurations are adjacent. In particular, suppose that a configuration is completely described by two integral-valued parameters x and X. Then the choice of these parameters would be such that passage from a configuration described by any pair of values (x, X) to one described by a pair (y, Y) is implausible in a short time unless either $|x-y|=1$ and $X = Y$ or $x = y$ and $|X-Y|=1$ or $x = y$ and $X = Y$ (in which case the original configuration is preserved). I shall now turn to an explicit example of such a system. I shall consider x to describe the number of units of an "uncapitalized" force operating in a region, and X to be the number of units of a "capitalized" force operating in the same region. The units of the uncapitalized force are all similar except for identity and position, and the units of the capitalized force are all similar except for identity and position (but the units of the two different forces may be dissimilar). Each unit is subject to annihilation either by virtue of acts of units of the other force or by virtue of natural causes independent of the presence of the other force. In order to simplify his illustrative model, I shall not allow for

birth or regeneration processes. The probabilistic nature of the death process may be described in many different ways depending on the application, and I shall here take one of the simplest possible. This process depends on four parameters, a, b, A, and B. Their general significance is listed below:

a and A represent expected rates of annihilation of individual units of the uncapitalized and capitalized forces, respectively, through causes independent of the presence of the other force;

b and B represent expected additive rates of attrition of the uncapitalized and capitalized forces, respectively, through the action of each unit of the opposing force.

Explicitly, let t be some time and h a short positive increment of time (shortness will be explained more precisely below), and tabulate the probability that there will be exactly (x, X) survivors at time t+h under various conditions which might prevail at time t.

If $x > 0$ and $X > 0$ and if there are (x, X) survivors at time t, then the probability of (x, X) survivors at time t+h is approximately $1 - (ax + bX + AX + Bx)h$;

If $X \geq 0$ and if there are (O, X) survivors at time t, then the probability of $(0, X)$ survivors at time t+h is approximately $1 - AXh$, and if $x \geq 0$ and if there are $(x, 0)$ survivors at time t, then the probability of $(x, 0)$ survivors at time t+h is approximately $1 - axh$;

If $x \geq 0$ and $X \geq 0$ and if there are $(x+1, X)$ survivors at time t, then the probability of (x, X) survivors at time t+h is approximately $[a(x+1) + bX]h$;

If $x \geq 0$ and $X \geq 0$ and if there are $(x, X+1)$ survivors at time t, then the probability of (x, X) survivors at time t+h is approximately $[A(X+1) + Bx]h$;

For any other populations of survivors at time t, the probability of (x, X) survivors at time t+h is approximately zero.

In all the above statements, the approximation is so good for small values of h that the ratio of the error to h approaches zero as h approaches zero.

Before these statements are translated into a set of differential equations, a few words of explanation might help the less mathematically inclined reader. The sense of all the statements including the last is that during a short enough time period there is essentially zero probability of more than one death. In the first statement the probability of everyone surviving is one minus the probability that someone is killed; the expected rate of deaths is given by the expression in parentheses which contains, in order, terms giving the rate of natural death of the uncapitalized force, rate of induced death of this force, rate of natural death of the other force, and the rate of induced death of this other force. This total rate is multiplied by the length of the time period to determine the probability of a single death during the time period, and the statement shown results. In the second statement, the first statement is modified by removing the term which would give a rate of induced death to a zero force, for in this case there is no potential victim. The third and fourth statements are developed similarly using rates for the attrition required to lead to the (x, X) configuration whose probability is being estimated. Probabilities related to the force which must not lose a unit if the (x, X) configuration is to be attained may be neglected (it turns out) in the approximation desired; specifically, such a probability is one minus a rate multiplying h, and this term in h has been omitted since it would multiply the term in h already computed to give a term in h^2, which may be neglected in this approximation.

These statements may be translated into a set of differential equations in probabilities of survival in two steps. These steps are first to write approximate conditional probabilities of survival equivalent to the above approximations and then to pass to an accurate limit depending on the last statement above concerning the validity of the approximation.

If the symbol $P_{xX}(t)$ is used to denote the probability of (x, X) survivors at time t, then in terms of conditional probabilities [5, Chap. V], the statements above may be rewritten for $x > 0$ and $X > 0$,

$$P_{xX}(t+h) = [a(x+1) + bX]hP_{x+1\ X} + [A(X+1) + Bx]hP_{x\ X+1} + [1 - (ax + bX + AX + Bx)h]P_{xX},$$

where the probabilities in the right member are all evaluated at time t, and the equations hold in the usual approximate sense.

Expanding the last term of this equation, placing the only term not involving h in the left member, dividing by h, and passing to the limit gives:

$$(1) \qquad dP_{xX}/dt = [a(x+1) + bX]P_{x+1\ X} + [A(X+1) + Bx]P_{xX} + 1 - [ax + bX + AX + Bx]P_{xX} \text{ for } x, X > 0,$$

and similar equations with the modifications in the last term indicated above for the case $x = 0$ or $X = 0$.

If there were initially (n, N) units assigned to the region, then a differential equation problem can be written to give the probabilistic result of the operation. This is the equation (1) suitably augmented by the cases for complete depletion and with initial conditions that, for all (x, X), $P_{xX}(0) = 0$ except that $P_{nN}(0) = 1$.

This gives a satisfactory formulation of this particular attrition problem for theoretical purposes. This type of argument may be extended to a wide range of problems. An argument not much different from this leads to a questionable approximation in terms of the Lanchester equations [6, pp. 63-77], and Neyman, Park and Scott have used a similar argument to study the survival of weevils [7]. I. J. Good and the author used the same type of development in an unpublished study in information theory, where the surviving units were messages consistent with all the information received.

Practically, however, there is much to be done before this model can be successfully applied to an attritive process. The parameters a, A, b, and B must be determined realistically, the model must be checked for suitability to the reality to be studied, and the differential equation problem must be solved in a form which will permit rapid calculation of numerical solutions. The particular equations developed in the example can be solved in closed form, but this form is complicated and of slight use computationally. In any event, there are $(n+1)(N+1)$ probabilities to be calculated, and this is an inconveniently large number if n and N are at all sizable; for example, if n and N are of the order of magnitude of a hundred, there are ten thousand probabilities at each time.

This completes the description of the example of a calculation of probabilities of results of actions which might be undertaken by the competitors—factor 1 above. I shall return to the question of easing the computational requirements in a later section.

I now turn to factor 2 above, calculation of utilities. Much of the basic treatment of this type of argument will be found in the texts by G. C. Evans [8] and P. A. Samuelson [9]. However, in the present short treatment I shall not follow any very complete exposition.

I shall here assume (as an example) a utility function which is additive over time. A single commodity will suffice for the example, and this example will depict a single holder of

this commodity with a competition between the utility, which will accrue if he expends some of his holding, and that which will accrue if he retains much of his holdings (as a safety allowance for unforseen future events, for example). Again, as before let h be a short time period, and let p represent the amount of the commodity on hand during this time period. Finally, let q denote the rate of expenditure of the commodity during this time period. The period will be assumed to be so short that the quantities specified do not change materially during the period. Then it is reasonable to think of a function U(p, q) which, with the usual approximation, leads to the utility accruing during the time period in question through the formula $J_t^{t+h} = h\,U(p, q)$. Here the left member represents the utility accruing during the time period from t to t+h, and the factors of the right member have been defined above. If the utilities are additive through time, the total utility through any time period is obtained (after letting h approach zero) by the integral $J = \int U(p, q)\,dt$ where the integration is taken during the time period in question, and the quantity p and the rate of use q are written as functions of t corresponding to the schedule whose utility is being evaluated. The formula may be augmented by including terms for the salvage value of any material on hand at the end of the period, for perhaps a cost of material on hand at the beginning of the period, and for any material received during the period.

With little imagination the field of application of methods such as this is seen to be exceedingly broad, so far as the underlying theory is concerned. The effective application to reality, however, involves a confirmation of the additive nature of the utility and, usually more serious, an explicit statement of a satisfactory utility function U. It seems reasonable to state that the competence which experience builds up in a talented individual operating in a decision-making capacity is equivalent to a competence in estimating and maximizing utilities. To the extent that a successful and experienced decision-maker could communicate an explicit formula for computing utilities, he could will to his successors his formula for operations. This type of communication would certainly be an immense aid in training and research.

If the class of admissible schedules is reasonably wide in a problem giving rise to a utility integral like the one above, the problem of determining the best feasible schedule may become one in the classical calculus of variations; in this case the numerical answer is likely to be difficult to evaluate for any reasonable utility function. If this allowable variation of the scheduling becomes at all complicated, the problem becomes one of dynamic programming, and the difficulties of computation are likely to be even more severe. These calculations really belong to factor 3 above, the determination of an optimal course of action. An artificial example of the type of problem which leads to a problem in the classical calculus of variations was treated by me in [10], but I shall turn to a problem more closely related to dynamic programming, described by me in [11], for an example of the involved nature which determinations of optimal courses of action may take. For an exposition of dynamic programming, see [12].

For the illustration, consider a competitive situation with the following conditions fulfilled:

1. There are two competitors, and the competition continues through a series of moves terminating at some indefinite time;

2. At each move each competitor must choose between one of several possible tactics;

3. There is a definite set of conditions under which the competition is terminated, and these conditions are certain to be attained within an unknown finite number of moves;

4. At the termination of the competition the value to each competitor of the configuration existing at the time of termination can be determined by means of a formula presented at the beginning of competition, and any gain by one competitor is countered by an exactly equal loss to the other;

5. At the beginning of each move one of a finite number of configurations must exist, and these configurations are partially order so that to each configuration there is a set of configurations (including the given configuration itself), which are no further from a state of complete attrition than is the given configuration;

6. The tactics available to the competitors depend on the configuration prevailing at the start of the move;

7. The result of any move is to replace the configuration prevailing at the beginning of the move by one no further from complete attrition, but, even with full knowledge in advance of the tactics to be used by both competitors, the resulting configuration may be estimated only probabilistically.

The general problem to be examined here is that of determining the value to each competitor of a particular inital configuration and of determining the proper tactics to attain that

As a more explicit example, let the configurations of two competitors (again uncapitalized and capitalized) be defined by the numbers (m, M) of units (similar for each of the competitors) available at the beginning of any move. Let the tactics available be the assignment of numbers (n, N) of these units to a region of active competition, any remaining units staying in a perfectly safe haven. Let the rule for termination be that the competition terminates at any move for which one of the competitors makes a zero assignment to active competition. Let the rule for progression from the configuration existing at one move to that existing at the next be determined by a probabilistic-attrition formula obtained by integrating the differential equation problem (1) above for a time period corresponding to the duration of a single play with initial conditions $P_{nN}(0) = 1$. Let the partial ordering be such that complete attrition occurs when neither competitor has any surviving units and let it be such that a configuration λ is no further from complete attrition than a configuration λ if, and only if, neither of the competitors has more available units in configuration λ than he does in λ. Finally, let the rule of terminal value be that the value to the uncapitalized player at termination is an amount dependent on the cause of termination (that is, whether the uncapitalized competitor alone, the capitalized competitor alone, or both competitors assigned zero units to the active region on the terminating play), plus an amount v for every surviving uncapitalized unit, minus an amount V for every surviving capitalized unit. If this value to the uncapitalized competitor is called u, then let $U = -u$ be the value to the capitalized competitor. Let no nonterminal value have any utility in itself during the move, so that all values are determined by the expected value at termination.

This particular example satisfies the conditions (1 through 7) set above.

In the application of a model such as the one above, the difficulties which must occur include both the assignment of realistic values to parameters and the computational difficulty of arriving at a correct solution. The necessity for, and difficulties of assignment of realistic values to the parameters are standard throughout any social science, and nothing further need be said about that here. However, there seems to be some point in outlining a straightforward computational attack on the problem and in estimating its extent.

The most obvious method of solution to this problem lies in an induction on the number of survivors. If the value has been computed for every configuration no further from complete attrition than is a particular configuration under consideration, then the value for that configuration may be computed by a difficult game solution.

An outline to a solution follows: Let the value of the configuration (m, M) be of interest. Suppose the value of every configuration (k, K) with $k \leq m$ and $K \leq M$ except for the configura- (m, M) itself is known. Let (n, N) represent tactics which might be chosen by the competitors; that is, $n \leq m$ and $N \leq M$ are numbers of units which might be assigned to the active region. The attrition rules give a definite probability of survival of every configuration in which the number of units surviving is not increased for either competitor. In addition, the number of uncapitalized units cannot be reduced by more than n and the number of capitalized units cannot be reduced by more than N, if the assignment (n, N) is actually made. The values of all but one of these configurations — that in which there is no attrition — is known. Thus, if the value of the (m, M) configuration is $u(m, M)$ then the expected value to be attained by the (n, N) assignment may be computed in the form $u_{mMnN} = r_{mMnN} + s_{nN}u(m, M)$. Here r represents the expected value if any form of attrition occurs (and it is assumed that the proper assignments will be made on future moves) and s represents the probability that there will be no attrition. It is not assumed that the (n, N) assignment is good for either competitor.

If it is assumed that $u(m, M)$ is known, then the expression above for $\bar{u}$ may be evaluated into a single number for each admissible (n, N) choice. If this is the case, it follows immediately from the theory of games (see [13] for example) that the proper assignments must be computed probabilistically: the probability of an uncapitalized assignment n is x_n and of a capitalized assignment N is X_N, where the numbers x_n are chosen to be non-negative, to sum to one, and to maximize the function $[\min_N (\Sigma_n x_n \bar{u}_{mMnN})]$; and similarly the numbers X_N are to be chosen to be non-negative, to sum to one, and to minimize the function $[\max_N (\Sigma_N X_N \bar{u}_{mMnN})]$. The maximum and minimum values computed above are the most favorably expected values from their individual points of view which can be guaranteed by the competitors, and the values (according to the fundamental theorem of the theory of games) are equal. Thus, the numbers x_n and X_N give the probabilities with which the various assignments should be made independently by the two competitors if the correct value of $\bar{u}$ has been chosen.

There is an easy test for determining whether the correct value of $\bar{u}$ was chosen, and an indication of the proper direction of change, if an incorrect value was chosen. If the proper value has been chosen, the value of the quantities calculated above will be exactly the same number $\bar{u}$; if the number chosen was too low the value computed will be higher than $\bar{u}$, and if the number chosen was too high the value computed will be lower than $\bar{u}$. The reasoning here is simple: it depends on the fact that the value $u(m, M)$ is precisely the highest expected value which the uncapitalized competitor can guarantee himself, and it cannot change if good tactics

are used. The existence of this value and the rule above about correcting bad guesses of the value are established by slight variations of a theorem and its proof by von Neumann [14].

The details of this calculation are not important here. What is important is that the calculation of the value of any configuration depends on the value of every configuration nearer complete attrition and that calculation of each of these values involves iterations and corrections based on calculations of the strategies and values of games—so that each step is a sizable one.

The mathematically interesting cases of dynamic programming are similar to the one described above except that the number of configurations becomes infinite, and inductive processes must be replaced by processes which are more similar to those of the calculus of variations. If the intermediate configurations have incidental utility of their own, extra complications are introduced. However, the purpose of this section was to introduce reasonably simple models which are themselves impressively demanding in measurement and in computation and to suggest an empirical attack on the problems thus raised. There is clearly no need to seek the most complex theories yet presented; the simplest offer sufficient challenge.

SIMPLIFICATION BY NUMERICAL APPROXIMATION (E.G., SCALING ATTRITION EQUATIONS)

In the last section I presented examples of difficult evaluations of parameters and difficult calculations which may turn up at many plausible places in the quantitative study of social sciences. I shall devote the rest of this paper to an examination of ways in which these difficulties may plausibly be alleviated. These ways fall into two categories—mathematical and empirical. Of these the simpler, from an expository point of view and from the point of view of justification, are the mathematical methods, and the more difficult are the empirical methods. The empirical methods are also those which require a professional understanding of the reality being considered and which will, therefore, be more interesting to the social scientist as he uses them successfully.

In this section I shall examine an attrition problem developed above. My purpose will be to indicate some purely mathematical methods, well-known to numerical analysts, whereby approximate solutions to these very extensive problems may be obtained with some considerable decrease in the amount of computation required. However, these solutions will be of no use in application unless the parameters can be determined with sufficient accuracy for the purpose of the application.

I now turn to the differential-equation problem (1) above. I shall consider the case where n and N are large so that the set of equations is too extensive to be attractive computationally. I shall make three observations:

1. For values of (x, X) other than zero, these equations may be rewritten in a form approximating a partial differential equation.
2. With careful handling of the initial value assignment, the solution of a partial differential problem based on the partial differential equation mentioned above will approximate the solution to the ordinary differential problem originally presented in (1).
3. A method of Budak [15] may be applied to write an initial value problem involving a smaller set of ordinary differential equations, and the solution of

this problem approximates that of the partial differential-equation problem and hence, that of the original problem.

The generality of this approach is not as great as would be desired. All reasonable attrition problems of the general type described lead to partial differential-equation problems which are analogous by straightforward application of classical methods elegantly developed by Feller in the works cited and elsewhere [1], [2], [3], and [4]. Budak's method may be used inversely to estimate the degree to which the solution to the partial equation problem approximates the solution of the original problem. In addition, his method can be applied directly to the partial differential-equation problem to get an ordinary differential-equation problem of reduced extent, whose solution approximates that of the partial differential-equation problem. However, there may still be some difficulty in computing the probabilities for configurations in which all the units of one of the competitors have been annihilated, and the error analysis for each problem seems to require separate attention (although the techniques are well illustrated in existing literature and simply need to be copied).

A formal conversion of the ordinary differential equations to a partial equation may be carried out as follows: Replace the variable x by a variable z related to x by the formula $x = nz$. Then z varies from 0 to 1 as x varies from 0 to n, and the size of z increments is $e = 1/n$ corresponding to increments of 1 in x. Similarly, replace X by Z with $X = NZ$. Z also varies from 0 to 1, and its increments have size $E = 1/N$. It is important in the argument that both e and E be considered to be small, for difference-quotients based on increments of this size will be replaced by derivatives. It will also be convenient to write $n = N\lambda$ and $E = e\lambda$. Finally, consider a function Q defined for all admissible values of (z, Z) and interpolated to intervening values, with $Q(z, Z) = PxX$ where PxX at any time t is the probability function of (2.1) and where (x, X) and (z, Z) are supposed to be related values under the transformation defined above. Note that a third argument, t, suppressed above, occurs in Q.

In the new notation, the equations of (1) not involving probabilities of zero survivors on either side may be rewritten:

$$\frac{\partial Q(z, Z)}{\partial t} = a\,\frac{(z+e)\cdot Q(z+e, Z) - z\cdot Q(z, Z)}{e} + \bar{b}\cdot Z\,\frac{Q(z+e, Z) - Q(z, Z)}{e}$$

$$= A\,\frac{(Z+E)\cdot Q(z, Z+E) - Z\cdot Q(z, Z)}{E} + \bar{B}\cdot z\,\frac{Q(z, Z+E) = Q(z, Z)}{E},$$

where $\bar{b} = b/\lambda$ and $\bar{B} = B\lambda$. If difference-quotients are replaced formally by partial derivatives, this expression leads to:

$$\frac{\partial Q(z, Z)}{\partial t} = a\,\frac{\partial[z\cdot Q(z, Z)]}{\partial z} + \bar{b}\quad Z\,\frac{\partial Q(z, Z)}{\partial zz}$$

$$+ A\,\frac{\partial[Z\cdot Q(z, Z)]}{\partial Z} + \bar{B}\cdot z\,\frac{\partial Q(z, Z)}{\partial Z}.$$

This is not quite the correct problem, for the function Q is not properly scaled as a probability function. The change arises from the fact that the functions P were probability functions over a discrete distribution but that it is desired to use the function Q as a distribution function over a continuous distribution. The most reasonable scaling is one which will cause the probability of a configuration to be contained in a particular (z, Z) region to be the integral of the scaled function Q over that region, integrating with respect to area: that is,

let $R(z, Z)$ be the function obtained by scaling Q properly, then the probability of having a configuration in some region of the (z, Z) plane is just $\int R(z, Z)\, dz\, dZ$, where the integral is taken over the region in question.

This scaling accentuates a difficulty which would have been present in any event—the provision of initial conditions such that the solution of the resulting partial differential-equation problem approximates that of the initial ordinary differential-equation problem. In particular, the initial conditions should be set up so that near the time $t = 0$ all non-zero values of R tend to concentrate near the argument values $(1, 1)$. This means more explicitly that $\lim [\int R(z, Z)\, dz\, dZ] = 1$ and $\lim [R(z, Z)] = 0$ for all (z, Z) except $(1, 1)$. In each case the limit is taken as t approaches zero from positive values, and the integral is taken over the whole region $0 < z, Z < 1$.

This is a standard kind of partial differential-equation problem in probability theory, and there is reason to believe that its solution would approximate that of the ordinary differential-equation problem. However, the existing proof of Budak limits the error in the approximate solution by an expression which depends on the first few derivatives of the solution of the partial differential-equation problem. These solutions tend to infinity under the formulation indicated, and it would seem desirable to modify the approach suitably to avoid low-order derivatives which are large. This is not a particularly demanding task, for it is possible to make fairly accurate estimates of the solution of the original ordinary differential-equation problem for short times, and indeed for times up to those for which zero inventories of units become probable. Integration over these times permits substitution of smoother initial conditions to the partial differential-equation problem by setting up the initial conditions to pertain at the time of termination of the integration of the ordinary differential-equation problem.

Finally, a difficulty connected with the change of the nature of the ordinary differential equation when forces with zero units are considered must be alleviated or avoided. This difficulty is avoided in the present approach, and a result is a lack of knowledge of the avoided probabilities. At least some of this knowledge can be restored by further involved calculation.

The avoidance of this difficulty is possible by a simple observation that, in the degenerative system set up, no probability of any configuration is affected by a probability of a configuration in which either competitor has fewer units than in the given configuration. Explicitly, the probabilities of configurations, for which either competitor has zero units, will not affect the probabilities of configurations for which each competitor has a positive number of units. Hence, these last probabilities may be computed without regard to the former, and if the last step, in which one competitor has no units, is deemed unimportant no difficulty ensues.

Finally, since the partial-differential equation is hard to solve, it is replaced by a discrete problem of some kind. Budak recommends a precise inverse of the procedure outlined above, and this is certainly reasonable. Other discrete methods, which involve taking discrete steps in time as well as in the two (z, Z) dimensions are also applicable.

Thus, there is outlined a purely numerical method of alleviating the computational difficulties which might be connected with the exploitation of any particular set of attrition equations. So far as I know, this complete process has never been carried out for any set of such equations, for there has never been a demand for this analysis. The field of application has been too limited to encourage the calculation and analysis involved, and it will stay so limited until tests for realism of the model and until means for evaluating coefficients involved

are developed. However, the analysis offers no insurmountable problem, and there should be little difficulty in arriving at a computational approximation of any attrition function satisfactory for most applications in the near future.

DESCRIPTION OF AN ALGEBRA AND ANALOGOUS FORMAL MODELS IN SOCIAL SCIENCES

Possibly the most easily described general mathematical model of wide practical applicability is an algebra. I shall present a short formal description of an algebra below, with the intention of presenting an analogous description of a formality whereby models of social phenomena may be set up and analyzed. My description of an algebra here is reasonably standard, but it might be appropriate to note that I follow Dilworth [16]. Somewhat similar approaches are to be found in many places, such as Albert [17].

An algebra consists of:

A system $\underline{A}$ of elements;
A set $\underline{O}$ of operations on these elements;
A set $\underline{R}$ of relations between elements;
A set $\underline{P}$ of postulates.

Strictly, the operations O must be finitary; that is, an operation is a rule mapping any set of a fixed finite number of elements to some element of the system. For ordinary arithmetic on integers, there may be several such operations: one maps an element to its negative, another maps two elements to their product, and a third maps two elements to their sum. The relations in this example may be relations of equality and of inequality. These are relations between pairs of elements, and a relation has the property that it is either true or false with respect to any set of arguments. The postulates are used to define the properties which are demanded from the operations. Thus, associativity, distributivity and commutivity as appropriate to arithmetical operations would be included in the postulates. In particular, a postulate states that a particular relation is true for the results of two different operations; for example: $a(b+c) = ab + ac$.

An algebra may be set up with any reasonable system of elements and sets of operations, relations, and postulates. Most of the ones which have been set up recently are complicated, so that study of the algebra requires considerable ingenuity. Techniques which have been developed for this include the study of sub-algebras, when they exist, or study of the algebra by means of homomorphisms (which will be described below) to simpler algebras.

A homomorphism of an algebra (A_1, R_1, O_1, P_1) to an algebra (A_2, R_2, O_2, P_2) is a mapping which, in an explicit sense, preserves all properties whose preservation in the second algebra is meaningful. As an example, the arithmetic of integers mentioned above may be mapped by a homomorphism onto an algebra whose elements are O and 1 by mapping even integers to O and odd integers to 1. In this new small algebra, the operation which is the image of addition maps (1, 1) to O and otherwise behaves like addition. The equality relation is preserved, but inequality in either sense in the larger arithmetic is mapped to a formal relation which is true for all pairs in the smaller algebra. The more explicit rules for this type of mapping are:

A mapping of a first algebra (A_1, R_1, O_1, P_1) to a second (A_2, R_2, O_2, P_2) is a homomorphism if, and only if, the image of the result of any operation of O_1 on elements of A_1 is the result of the image of the operation in O_2 applied to the images of the operands in A_2;

if truth of a relation of R_1 between elements of A_1 implies truth of the image relation in R_2 between the image elements in A_2 (but falseness does not necessarily guarantee falseness of the image relation), and if the images of the postulates in P_1 are precisely the images in the second algebra of the statements of these postulates as relations in R_1 between operations in O_1 on elements in A_1.

The point of using homomorphisms to study algebras is the same as the point of using any classification scheme to describe a complicated system by a simpler one: it is hoped that useful parts of the structure of the complicated system may be retained and observed or studied in the simpler system. (This is the point of the index to a book, for example; the index does not tell what the author says about a particular subject, but a good index properly used will indicate whether the author discusses the subject at all.)

My purpose for the superficial description of algebras is to suggest that the same type of factorization and description of the processes involved in quantitative studies of social sciences might serve as valuable a purpose in the development of these concepts as has the description of algebras in illuminating the aims and processes of that branch of mathematics.

In particular, a quantitative social science might be based on three disjoint systems of elements: sets of inventory levels, distribution functions over the population of sets of inventory levels, and distribution functions over the population of these distribution functions. These three systems of elements will be denoted respectively by: A_i, A_d, and A_D.

The elements of A_i are ordered sets of numbers representing the states of the various inventories which make up the reality considered, or representing the states of the various inventories which make up a simplified abstraction of this reality. Thus, if an abstraction requires ten thousand numbers to describe the numbers of various units available and their assignments, the elements of A_i in this model would consist of sets of ten thousand numbers consistent with the conditions imposed on the model. More specifically, the total inventory of a commodity, such as the number of male students in a school, for example, might be designated as one inventory (hence as one of the numbers in an element of A_i); and the number of male students assigned to the sixth grade might be designated as another inventory. Consistency would demand, among other things, that the number of male students assigned to the sixth grade not exceed the number of male students in the school. Generally speaking, an element of A_i must specify not only the available level of each important commodity, but also the levels of this commodity assigned to various places or uses.

It is reasonable to assume that the elements of A_i are observable and can be measured.

The elements of A_d are functions whose arguments are elements of A_i and which are probability distributions: that is, each element of A_d assigns a nonnegative real number to each element of A_i. If the number of elements in A_i is finite, the sum of these numbers is one; and if the number of elements in A_i is infinite, some generalized sum (an infinite sum, or an integral, or a combination of the two) is still one. Each element of A_d attaches to eacn element of A_i a probability that that element of A_i describes the inventory configuration existing. Many of these assigned probabilities would normally be zero.

A discussion of A_D must follow closely the themes of the discussion of A_d. In short, each element of A_D will give a probability that any element of A_d properly describes the distribution over the admissible inventory configuration. The role to be suggested for the system A_D will become clearer later.

It must be noted here that the total system of elements A_i, A_d, and A_D is not homogeneous, as it was for an algebra, but the elements can and should be well described.

It should also be noted that, while elements of A_i are observable in the sense of being directly measurable, elements of A_d and A_D are not observable and hence, are only indirectly and imperfectly measurable.

The set of operations is more revealing. There is one reaction operation: mapping A_i to A_d. That is, if a model of successive moves is assumed, each inventory configuration at the beginning of a move determines a probabilistic distribution of inventory configurations at the end of the move. The change in inventory configurations is caused by an interaction between inventories. Since the elements of A_d are not observable, this reaction operation is not directly measurable. Another operator, which will be crucial in the model suggested here, is a Bayes operator which maps the cartesian product of A_D and A_i to A_D. This means that there is a process which replaces any element of A_D by an element of A_D based on an observation of an element of A_i (an element of the cartesian product of A_D and A_i is an ordered pair of elements, the first from A_D and the second from A_i).

Finally, there is an operator mapping the set of true relations to A_D. Its nature will become clear when the relations are discussed.

The relations which are important in a social model are those upon which the mechanics of the reaction between inventories is based. These relations are the mechanistic connection between the model and the reality, or rather the abstract description of this relation. By relating various different objects, the objects are classified. In particular, this process may permit the scientist to predict the expected wear on, or effectiveness of, an entirely new item; basing his prediction on his knowledge of similar items which have been observed in the past. This prediction is common and indeed necessary, but it need not be made to appear to be made on a completely arbitrary basis. The most crucial part of the formal system, which is being expounded here, is setting up and utilizing relations, which lead to these estimates.

Before I develop the concept of relation to be presented here, I shall outline some differences between it and the relations of algebra. An algebraic relation is a precise statement. In arithmetic, it is a statement that one number is equal to another, or that one number is greater than another, or that one number is not smaller than another (this last admits the possibility of equality). The statement is perfectly clear, but it may be true or false depending on which number is chosen. At least implicitly, the algebra furnishes a means for testing the truth of the statement; thus, the statement that 5 is less than 3 is obviously false.

The relations proposed here for models of social science are not decidable in this manner. They are intended to furnish an agent for introducing intuition into the formulation of a model. In an imprecise way, it seems clear that objects which are similar in construction and similar in application, or which are subject to similar attrition or wear, are related in terms of a science in which these properties are important. However, if this science is to be presented precisely, a more precise instrument must be furnished for describing this relationship.

Algebraic postulates are stated in a form which asserts the truth of a relation between a pair of operators acting on an ordered set of operands. Thus, the statement $a(b+c) = ab + ac$ may be taken as a postulate (or, if more elementary postulates have been assumed,

as a theorem) of ordinary arithmetic (and of many other algebras). The objective in this exposition is to furnish some similar logical model applicable to the social sciences. This model should have following properties:

1. It must furnish a means for precise statements of postulates.
2. It must furnish a means whereby conclusions from a postulate, based on the truth value of a set of relations, may be modified in the light of experience without contradicting the postulate.
3. It must offer an opportunity to test the validity of a postulate through experiments involving observable elements.

The system being outlined satisfies these conditions, and the present task in the exposition is to introduce relationships and to formalize the structure of postulates in a way which will attain the goals.

To this end, consider a set of unitary relations; in particular, for each object which is counted in any of the inventories whose levels constitute an element of A_i, the relation is either true or false. The relation, in fact, states a property of the object in question. The more relations included in a theory in an essential way, the richer and the more detailed the theory; the fewer relations included, the simpler and more easily checked the theory. A natural development gives a somewhat broader class of relationships—"similar but not equivalent" relations may themselves be treated as special cases of a class of relations . Such treatment may simplify the realistic application of a theory, but it complicates the first presentation of the formality of a theory.

The essential feature of the present theory lies in the structure of a postulate. A postulate is a mapping of the set of truth values of the relations of a theory to the class A_D.

This means that a postulate is a statement concerning the interactions of inventories of objects with particular sets of properties. (Formally, an object has a property if the relation stating that the object belongs to the class of objects having the stated property is true; the notation here is clumsy and unnecessary so long as only unitary relations are considered, but it is retained because of the analogy with algebra and because more involved natural developments are best presented in the guise of relations.) However, it should be noted that there is nothing in the system which describes in a deterministic way the reactions of inventories. The most precise statement arises through selection of an object from the set A_d. Such a choice gives a probabilistic description of the reaction. However, choice of an element of A_d as the image of a set of truth values would not permit further modification of a theory (short of discarding it) on the basis of experience, and it is for this reason that the set of objects A_D was introduced. An element of A_D is two steps removed from a deterministic statement of reaction (although for a single sample it can be used to determine a probabilistic-distribution function over A_i) in that it states that any element of a subset (which may or may not be the whole set) of A_d may, in fact, govern the probabilistic interaction and that the probability of any particular element being the one which governs is as stated by the chosen element of A_D.

Now, this two-level probability argument permits modification in the light of experience and rejection in the light of overwhelming statistics. This is accomplished by a slight modification of terminology. The postulates should be read to state that the element of A_D,

assigned as an image to a particular set of truth values of relations, is the a priori probability-distribution function over the population A_d. This means that among all objects in our experience satisfying the hypothesis of the postulate studied, the reaction-probabilities are distributed in A_d in accordance with the stated element of A_D. Further information concerning any identified object may change this distribution, so that it is expressed by some function which may not even belong to A_D. On the basis of experience, an a posteriori distribution may be calculated, then, and this calculation is to be based on the Bayes operator described above in the list of operations in this model. A specific formula for this operator will be presented in the next section.

There has been outlined above a general form for a theory in social sciences which might be properly responsive to experience. This exposition will be completed by a simple example in which there is a single inventory. The example is based on reported experience with gasket demands in the overhauling of aircraft engines. Each move will be the overhaul of one engine. The usual assumption is that the demand for gaskets of any particular type will average about 1.25 times the number in the engine. (The excessive number is based on the fact that every gasket is destroyed when the engine is taken down, a few new gaskets are ruined while installation is being attemped, a few are defective, and a few are lost.)

Two different formalisms, each an example of the logical structure outlined above, will be presented; I shall contend that the first is unrealistic and that the second is (presumably) realistic.

In each case, each element of A_i consists of a single element: the number of gaskets which must be used to meet the overhaul-demand for the engine being treated.

In the first case, the set A_d contains a single operator stating that each overhaul will generate an inventory (in the engine, or in the trash can, or in the lost category) of g gaskets where g is a number which is the sum of the number of gaskets in the engine and a number drawn randomly from a Poisson distribution (a Poisson is a classical probability-distribution popular in examples like this one) with mean equal to 25 percent of the number of gaskets in the engine.

In this first case, since A_d has only one element, the set A_D is trivial in nature: it must contain a single element which states that the probability of the single element of A_d governing the demand is one.

In the second case, the set A_d will contain an infinite number of elements. There will be one element for each real number m lying between .20 and .30; the element, corresponding to any number m, states that the demand inventory (as described above) will be g gaskets, where g is the sum of the number of gaskets in the engine and a number drawn randomly from a Poisson distribution whose mean is the product of m and the number of gaskets in the engine.

In this second case, the set A_D may have significance. It will consist of only one element, for there is only one set of truth values (an engine is an engine is a true statement). This element might assign a probability zero to every element in A_d, except the element for which $m = .25$; in this case, the formalism is entirely equivalent in effect to the formalism in the first case. However, any other probability-distribution over A_d might be chosen as the element of A_D, and the choice of such a distribution is the only postulate of the system. For illustration, choose the uniform distribution; that is, choose the following postulate:

POSTULATE if the element of A_d corresponding to m actually governs the action, and if a and b are two numbers with $.20 < a < b < .30$, then the a priori probability that m lies between a and b is 10(b-a).

(The number 10 appears in this statement because the length of the interval of allowed values of m is 1/10; specifically, if a = .20 and b = .30 the probability must be one.)

It requires little discussion to convince one that the first formalism is unrealistic. Kronnecker once remarked that, "The integers were made by God; all the rest is the work of man." However, God has shown an overwhelming preference for man's work in assigning values to quantities which have no binding reason for being integral, and it would be surprising indeed if any single number named in advance would be the right mean for the distribution selected. Enough care in counting replacements would serve to reject the first formalism, and the nature of the statements provide no possible modification.

On the other hand, the second statement, while undoubtedly wrong, may not be provably wrong during the life of the program. If this is the case, and if there is no evidence for a more plausible formalism, then it must be accepted at least as a theory which has not been disproved. After a very long time and many demands, the a posteriori probability may give great assurance that m must lie very near to some particular number in the range (or tests may indicate that the assumption that it lies in the range at all is implausible and that the theory should be rejected), and this first sharpening of the theory will be completely consistent with results.

The postulate, however, was based on experience with many nonidentical gaskets. Such a postulate (usually, unfortunately, in the first form rather than the second) is universally made in order to set reasonable stock levels. It depends in this case on long experience with an inexpensive item. In many other cases, the postulate itself is a result of long calculation based on assumptions of the mechanism which governs the interaction. The probabilistic nature of the reaction may be felt to be real, in that no amount of information would give a correct deterministic prediction of the interaction, or it may be felt to represent the lack of knowledge available to apply to the problem. In either case, a reasonable set of axioms governing the probabilistic approach may be introduced; for this see L. J. Savage [18] or a discussion by W. E. Smith [19].

FORMULAS OF FINITE BAYESIAN STATISTICS

Formulas pertinent to the formulation in the preceding section will be presented here essentially without discussion. (Formal expositions are available in several places, including [18] and [19].)

Let A_i (as above) be a finite set of plausible events, and let A_d be a finite set of probability-distributions over A_i. Finally, let A_D be a finite set of probability-distributions over A_d.

Choose a single element from A_D (through application of a postulate); call this element z. This element is a probability-distribution function over A_d and if y is any element of A_d, the value of this function for the argument value y will be denoted by Pr[y]. This is the probability that the distribution y governs random choices from the population A_i. Similarly, for any choice of y in A_d, the value of the corresponding distribution function over A_i (for a particular argument x which is an element of A_i) will be denoted by $Pr[x \mid y]$, the conditional probability that the sample x will be drawn from the population A_i

in a single trial if y is the governing distribution function. The function whose value for any element x of A_i is $\Sigma Pr[x|\bar{y}]Pr[\bar{y}]$, where the summation is over all $\bar{y}$ in A_d, is the unconditional probability of x (z being ignored); it is denoted by Pr[x]:

$$Pr[x] = \Sigma Pr[x\ \bar{y}].\ Pr[\bar{y}].$$

For each element x of A_i for which $Pr[x] \neq 0$ and for all elements y of A_d define:

$$Pr[y|x] = \frac{Pr[x|y].\ Pr[y]}{Pr[x]}\ .$$

Remarks:

Pr[y|x] is the a posteriori probability of y based on a sample x and the a priori probability Pr[y].

For each admissible x, Pr[y|x] is a probability-distribution function over A_d.

Intuitive meanings may be attached to all these formulas by ascribing to probabilities an interpretation, as frequences of average (independent) occurrence.

A very low value of Pr[x], where x is a sample which has actually been observed, implies low plausibility that Pr[y] correctly depicts the actual probability distribution function governing the choice of y (hence, it indicates that the original postulate should be reexamined or rejected).

CONCLUSIONS — A CASE FOR FORMAL EMPIRICISM

The preceding sections have been preliminary to suggestions concerning realistic treatment of quantitative aspects of social sciences. These suggestions, now to be presented, will be based on acceptance of some techniques used in the development of sciences which have been using quantitative theoretical methods longer and on rejection of other techniques of these sciences.

In particular, it should be noted that no precise theory is proved to be valid as a result of statistics based on actual experiment. In the long run, all these statistics can do is to indicate that a particular theory is implausible. Given a set of experimental results, there must be an infinite set of possible theories consistent with the results. Of these, the one which can be explained by mechanisms which are intuitively most acceptable or which are subjectively the simplest is accepted until the weight of anomalies forces its modification or rejection. This general program seems to be the only constructive one available.

On the other hand, there has always been a strong desire for universality in a theory. While this desire must be as strong in social sciences as it is in the physical sciences, it must not be allowed to inhibit or prevent the development of precise theories. Certainly the universality of Newton's laws of motion, applying equally well to the smallest particles he observed and to the moon and planets, was spectacular and pleasing. But the development of some aspects of the theory of heat and thermodynamics before successful attempts to base the theory on statistical mechanics was useful and essential in much of the development of physics.

The social scientists are presently coming into possession of quantities of data never before available, statistical techniques and knowledge which is considerably advanced over that available a few years ago, and data-handling equipment which is radically powerful. All these assets combine to make possible a fast development of these sciences along lines which were not available or required during similar periods of development of the physical sciences.

In particular, the variances encountered in measurable probabilistic phenomena in social sciences exceed anything physicists had to contend with prior to quantum mechanics; but techniques and equipment for handling such phenomena are now becoming available.

These techniques, especially when they require the use of automatic data-handling equipment, require great precision in expression; and some parts of this paper have considered this problem of precise statement.

Now the general picture to be faced is that, no matter how precisely and accurately the mechanics of a real situation is depicted, there seems to be reason for suspecting that predictions can be made only probabilistically. This is partly due to completely intrinsic reasons similar to those which have led to the same conclusion in quantum mechanics (in an attrition case, it might be conceivable that an exact lifetime of a piece could be estimated but only at the expense of destroying the piece and hence making it unavailable to serve this lifetime), and it is partly due to the fact that social reactions require communications and decisions; and the facilities for these functions are so limited that decisions must be made with incomplete information and with inadequate contemplation in order to meet a schedule imposed by nature or by associates (cooperative and competitive).

At the same time, no matter how precisely the general mechanics of a situation are known, useful application (to give these probabilistic results) requires an adequately accurate determination of parameters and a facility to carry out computations based on the model and the determined parameter values.

Each estimate of a parameter value in turn requires measurement, and, in general, increasing confidence (in the estimate of a parameter value) is gained only through increasing the number of measurements.

It has been noted above, in passing, that even for as precise a doctrine as algebra, there is little trouble in generating a system which is too complicated for complete analysis. In such a system, the postulates are not related to any realism and do not have to be tested (except for consistency); but even in such a system it is frequently necessary to map homomorphically to a simpler algebra in order to gain knowledge about some of the major characteristics of the more complicated algebra.

The analogue of this homomorphic simplification, in the formal description of a social model suggested above, lies in reducing the number of relations; that is, in reducing the number of properties considered. This reduces the number of postulates, and hence the number of parameters to be determined. A quantitative analysis of the following statement is too involved for inclusion here, but it seems intuitively obvious that, if each of a set of parameters has a value close to that of all the other parameters of the set, and if information is obtained about one parameter for each measurement made, it is economical to assume that all parameters have the same value and to apply any information obtained about one to evaluating all until enough individual information is available to permit meaningful estimates of the differences between the values of the individual parameters. Thus the analogue of the simplifying homomorphic mapping in algebra not only retains the advantage of reducing the computation to a complexity which may be met, but it adds to this advantage the advantage of permitting an earlier test of hypotheses and a more rapid a posteriori adjustment based on experience. Even if it is understood that the gain of this second effect is becoming nebulous (because of a large number of recorded measurements), it is still true that continuation of use of the simplified model has a cost only of making the probability distributions of the theory more diffuse,

and this might lead to negligible practical degrading of the information to be gained from the application of the theory. This last is particularly true if the application of the more accurate theory involves computations so extensive that they are not feasible during the period available in the application, so that approximations must be used anyhow.

Finally, it should be noted that the axiomatics of the Savage approach [18] are such that they readily admit the use of probability arguments based on incomplete information. Specifically, the a priori probability has the meaning of a conditional probability: in our experience it is plausible to conclude that an object having the listed properties, and with all other properties unknown for either intrinsic reasons or for pure lack of communications and information, gives rise to the element z of the set A_D. The lack of communications itself may be induced by a number of causes—from not opening mail because of laziness, to not opening mail because of a realization that the information contained will not realistically help the program, to not opening mail because of lack of time to read it, and on to not opening mail because it was not received. In any of these cases, the estimates of reactions to specific configurations can be made rationally on the basis of information available at the time of the estimate.

This argument leads to an argument that value itself may be treated as an inventory item, so that dynamic-programming situations can he handled in the same way. Thus, if some elementary knowledge of the general behavior of a value or a utility function is at hand, and if it is assumed that successful and experienced practitioners in some demanding field of endeavor have arrived at some implicit scheme for estimating values or good tactics, then their reactions may be used as experimental inputs for a posteriori adjustments. In short, if the extensive calculations of value indicated in the second section (Some Examples of Analyses to be Supported) are unacceptably or unrealistically extensive, the generation of value by configurations of other inventories may be estimated probabilistically using this same model, and the estimate may be adjusted or rejected on the basis of continuing experience.

A final word should be inserted concerning the computation involved in a posteriori adjustments. It is likely to be extensive, as Smith [19] has demonstrated. However, once the postulates of a theory are set, there is material in hand sufficient for the application of the theory, and if these postulates have been carefully set by the most competent scientist available for the job, this material is the best which can be available. Continuous adjustment is indicated until the information injected into the theory is so great that assimilation of additional information is too expensive to justify, in the light of its probable unimportant effect on the a posteriori probabilities which are being modified, or until the information indicates that the theory itself is implausible, so that there is an opportunity and a justification for development of a new theory.

In general, this means of injecting empirical knowledge into a precisely stated theory in the above formalism seems likely to be close to the implicit processes which are used by people in making decisions in competitive situations. A chess player, for example, makes some such implicit estimates. An advantage of the precise formalism suggested here is that the information gained through experience is readily transmitted to later generations of workers, and thus, a continuing improvement may be guaranteed. A disadvantage is that calculation by the human brain for fast reactions seems to be more effective if it is based on implicit models and on implicit reactions to input information, whether this information is implicit or

explicit in its nature. (I am not certain that this is a real effect rather than a matter of training. Many engineers seem instinctively to reach for slide rules and tables to draw quick conclusions which are frequently no more or less erroneous than those drawn by leaders in affairs involving social sciences who arrive at conclusions by gazing with half-closed eyes into clouds of smoke from their pipes.)

My conclusion is that rational development of a social science may frequently be based on the Bayes model I have outlined in this paper. There is no requirement for mechanical justification of any postulate, although a complete explanation of events through extension of the mechanisms governing simple reactions is a most desirable ultimate goal of any science. Each postulate is subject to statistical testing, and this test should be a normal part of the development of the theory. No statistical test establishes the validity of the postulate, but the model proposed invites a refinement of the information available on the basis of such statistical tests. At any time, the theory is applicable through the use of the probability distribution with value $Pr[x]$ introduced in the fifth section (Formulas of Finite Bayesian Statistics), if it is assumed that functions based on the probability functions and having importance in the theory are all linear, or linear within a tolerable error. Otherwise, an expected value of important functions based on the whole present estimate of the governing probability distribution over A_d may be calculated with considerably greater effort. In short, in many fields of application the use of such a formulation of existing knowledge would be practical, economical, and probably rewarding.

REFERENCES

[1] W. Feller, "Generalized second order differential operators and their lateral conditions," Illinois J. of Math, v. 1, 1957, pp. 459-504.

[2] W. Feller, "On the intrinsic form for second order differential operators," Illinois J. of Math, v. 2, 1958, pp. 1-18.

[3] W. Feller, "Differential operators with positive maximum property," Illinois J. of Math, v. 3, 1959, pp. 182-186.

[4] W. Feller, "On the equation of the vibrating string," J. of Math and Mech, v. 8, 1959, pp. 339-348.

[5] W. Feller, An introduction to probability theory and its applications, v. 1, second edition, John Wiley & Sons, New York, 1957.

[6] Philip M. Morse and George E. Kimball, Methods of operations research, The Technology Press, Cambridge, Mass. and John Wiley & Sons, New York, 1951

[7] Jerzy Neyman, Thomas Park, Elizabeth L. Scott, "Struggle for existence. The tribolium model: biological and statistical aspects," The Third Berkeley Symposium on Mathematical Statistics and Probability, Proceedings, v. 4, pp 41-79, U. of Calif. Press, Berkeley, 1956.

[8] G.C. Evans, Mathematical introduction to economics, McGraw-Hill, New York, 1930.

[9] Paul A. Samuelson, Foundations of economic analysis, Harvard Univ., 1947.

[10] C. Tompkins, "Allocation between assignments to planned consumption and to safety allowance," Logistics Papers circulated by The George Washington University Logistics Research Project, Issue 3, 1950.

[11] C. Tompkins, "Probabilistic problems and military evaluation, an example," Logistics Papers circulated by The George Washington University Logistics Research Project, Issue 4, 1950.

[12] Richard Bellman, Dynamic Programming, Princeton, 1957.

[13] J. D. Williams, The Compleat Strategist, McGraw-Hill, New York, 1954.

[14] John von Neumann, "A Model of general economic equilibrium," Rev econ studies, v. 13, 1945-6, pp. 1-9 (translation by G. Morgenstern from Erg. eines Math Kolloq, v. 8, 1937, pp. 73-83).

[15] B. M. Budak, "On the method of straight lines for certain boundary value problems," Vestnik Moskov. Univ., Ser Mat Meh Astr Fiz Him, v. 11, 1956, p. 3-12.

[16] R. B. Dilworth, unrecorded lecture, University of California, Los Angeles, 1960.

[17] A. A. Albert, Fundamental concepts of higher algebra, Chicago, 1956.

[18] L. J. Savage, The foundations of statistics, Wiley, New York, 1954.

[19] Wayne E. Smith, A posteriori probabilities, Dissertation, University of California, Los Angeles, 1958.

* * *

PANEL 1
HARDWARE FOR LOGISTICS RESEARCH

CONTROL OF AND COMMUNICATION WITH DATA-HANDLING MACHINES

William Orchard-Hays

C-E-I-R, Inc.

Change and reworking of systems has become a way of life in the computing or data-processing field, with nearly a decade of major advances. Both the hardware and the applications of the last few years can elicit only admiration. But mingled with this is a tinge of anxiety that hardware is getting ahead of techniques for utilizing it. Logistics research is a field which offers an extreme technical challenge here. If the equipment is not a responsive "member" of the team, it may be as much of an impediment as a help.

Ever larger programming groups as a means of getting the increasingly complex problems "on the air" are becoming more ineffective. There is a definite limit to the availability of expert programmers but, even when they are found, they can solve only one or two of many classes of problems involved. A few basic principles can be set forth to clarify this. Pre-eminent among these is the fact that the handling of information generates more information which must in turn be handled. This can continue ad infinitum unless restraints are imposed in the nature of policies which reflect the purpose at hand.

The over-all problem seems to be one of communication between people. A particular technical problem that arises is making the hardware responsive to the user's needs without an overburdening programming load. A scheme for accomplishing this is outlined. It will be demanding of the users but hopefully rewarding. Responsibility is a necessary counterpart of responsiveness.

As to the supposed dissimilarity between hardware for scientific and for data-processing applications, it is observed that the really hard problems for both are identical. As always, the primary problem is the proper organization and distribution of human effort.

For nearly a decade the computing, or data-processing, field has been witnessing such a series of major advances that change and reworking of systems has become a way of life. Some of the original enthusiasm for bigger and better projects to better automate computing and data-handling techniques has worn a bit thin. It appears sometimes that the equipment users have not kept pace with the equipment makers in imagination, sensible distribution of effort, and just plain pride of workmanship. Yet a realistic evaluation of the work done on large machines in the past three or four years will elicit admiration from anyone capable of

understanding it. But mingled with the understanding admiration is likely to be a trace of fear that man as yet cannot utilize effectively the powerful machines which he is building. Some of the most ambitious, thoroughly planned, and well-funded projects have been little short of fiascoes. And machines five times as powerful are on the threshold with other machines twenty times as powerful as that following.

Any one who has an interest in data processing should perhaps feel a tinge of this anxiety. There is a monumental task ahead of us, if these machines are to fulfill their promise, and we should recognize it. I can think of no application which offers a greater technical challenge than logistics research. The need to control and to communicate with data-handling machines is at a peak here, much more so than in so-called routine operations. If the data-processing equipment is not a responsive "member" of the team, it may be as much an impediment as an aide in the effort.

The standard solution to the problem of "getting on the air" with a big project has been to proliferate ever-larger programming groups—by hiring, training, renting, pirating, or any other means, including dumping a bunch of people in to sink or swim. This method is becoming increasingly ineffective. A large number of people doing a mediocre job only adds to management's problems. This is not to say that the people are mediocre. Rather, they often lack direction. However, a definite limit to the number of potentially top-flight professional programmers—dedicated, intensely interested, highly skilled people capable of creative work—is apparent to anyone in the industry. This limit is indeed small, I would guess in the hundreds. The reason for this is perhaps not so apparent. I submit that it is but a part of the over-all problem of automation and readily explainable when a few basic principles are understood. Let us begin with a few fundamental ones:

1. The handling, movement, and minipulation of data involves work, that is, the expenditure of energy with movement in space in finite time. The fact that the data is recorded in magnetic fluxes of some kind and moved as electronic pulses does not alter the truth of this principle.

Lest you think my observation trite, let me point out that a computer can be kept on all day, in perfect working order but doing nothing useful, and it costs just as much money as though it is carrying out the most complex calculations, except for the relatively trivial cost of tab paper used, etc. It is easy to conclude that, once the initial investment is made, the computer can be used in any way at no additional expense, so far as machine costs are concerned. While this may be true—it depends, of course, on your method of accounting—such an attitude is fatal to efficient and effective use of the hardware. It is just as wasteful to do things in a roundabout or stupid way on a computer as to send out a fleet of delivery trucks with randomly distributed destinations. When a 2-million-dollar machine is used to do something, it should be purposeful, with an expected end result.

2. Since the handling of data involves the generation and manipulation of information about information, the growth of data-handling requirements is self-compounding unless carefully constrained.

This principle is insidious, but true. It might be called the Parkinson's Law of data processing. It is interesting to note that human beings, in doing clerical work or calculations, impose the necessary constraints in the form of what we call common sense and judgment. The dullest sort of person is able to exhibit these qualities in far greater measure than we often admit, but the brightest genius has difficulty in precisely defining them—which leads to the third principle.

3. Nothing can be mechanized which cannot be precisely defined. This is as true of second, third, and higher orders of control as of primary mechanization.

I have come to the conclusion that many people who ought to know better don't really believe this principle. Sometimes I try to ignore it myself, always to my eventual grief. It is too tempting to solve the easy problems, pushing the hard ones back with a sort of mental note that "we'll do something" when the time comes. If pressed we may mumble something about statistical methods, which sounds impressive. Now I do not want to appear to minimize the worth of faith, self-confidence, and drive. But it is foolish to undertake a piece of work without having first estimated the effort required. In data processing, we have sometimes been trapped into doing this because of a failure to understand the true complexity of what we are asking for. Thus complex, far-reaching, and important decisions and plans are left to those who are unqualified to make them, technically, administratively, or policywise.

The inability to define precisely all aspects of a technology does not bother us too greatly in most endeavors. A car still requires a driver; a plane, a pilot. We think of an assembly line as highly mechanized, but it works only because many people are working on it and making little, and sometimes big, spot judgments all the time. But who can work at the rate of 40,000 additions per second, even in a supervisory capacity? Of course, the very essence of the stored-program concept is to allow complete preplanning, which amounts to canned supervision, of the whole sequence of operations required by some process. There are two catches to this: The first is our second principle, i.e., the bookkeeping soon outweighs the data-processing operations and becomes a bigger data-processing problem than the original one. The second catch is that it is impossible to preplan the proper action for every eventuality; in fact, it is questionable whether the term "proper" has any meaning in this context except as dictated by particular policy. All this is simply a roundabout way of saying that human beings cannot be taken out of the daily activities for which they are responsible and which exist in the first place by their decision. The introduction of new machinery—no matter how revolutionary, powerful, or potentially useful—does not change this basic fact of life. Rather than relieving people of effort or requiring less in the way of disciplined endeavors, this new equipment is more demanding if its use is to be profitable.

Let us admit boldly, or humbly if you prefer, that there is a lot of confusion in data-processing work today. Let us further admit that there is a great deal of difficulty and uncertainty in the projects to which DP equipment is often applied. If this weren't true, most of us would be out of work. Now, while it may be possible to cut down on the confusion, it appears realistic to me to assume that the problem-difficulties will increase, certainly not decrease. These problems are not the exclusive domain of any one discipline. Certainly they do not all belong to the field of programming. Programming has its share of tough problems and also its share of skill and helpful techniques to be contributed. What makes the programmer invaluable is that he knows how to make a computer perform. What makes him inadequate is that he doesn't always have the script. After all, problems originate in people's minds, and techniques of operation reside in other people's minds. We need to remember, and to believe, that the really valuable residue of a completed job consists of the ideas, insights, and techniques recorded in human brains—in short, the experience gained. The real problem confronting us is in communicating between the variety of skills used on today's large projects.

Communication evidently means different things to different people—which seems odd. But one must recognize that vast efforts are underway to increase communication between

systems, subsystems, or just plain code and systems. It is not clear to me that communication between two inanimate objects can exist, but I admit my knowledge of psychology is limited. At any rate, I am more concerned with communication between humans and between humans and machines.

Now there are at least three important conditions for communication:

1. The parties concerned must have rapport in the area under discussion;
2. The quantity and rate of information sent must be within the capacity of the receiver to assimilate;
3. The language and medium used must be known and compatible to both parties.

A lack of any of these can be compensated for by sufficient determination, repetition, patience, and time. In EDP, however, efficiency must be given great weight. We cannot always afford to compensate for poor communication.

It might be asked how one can gain rapport with a machine? The best way is to build into the programs the kind of receptors and output devices that one would like. But this option is usually available only to a limited number of people, namely, expert system programmers. But once the user is one step removed from the machine, as say a production programmer, the communication is mainly between people. Furthermore, the system-programmer now has a new responsibility, viz, to provide techniques which can be responsive to the user without major overhauls in handwritten code.

Turning to the other end of the spectrum, how can the language and working papers of the physicist, economist, engineer, or logistics researcher be compatible with the language and experience of the professional programmer? Except in isolated cases of fortuitous backgrounds, the answer is they can't. The truth is that there is a whole spectrum of languages from the problem formulator to the machine operator, with only slight intersections. I dislike to call these language "levels." It is not a question, at least not anymore, of one being above or below another. The language of the engineer is every bit as specialized as that of the programmer. Generalizations can be made in both directions, but this may or may not help to get a job "on the air."

Now the machines are not the exclusive domain of any one group, any more than the problems, except in an administrative sense. The problem formulator is just as anxious, and probably more needful, to exercise control and introduce changes, from his viewpoint and with his language, as is the system programmer. I think we all recognize the fact that human beings still have to work together in formulating the roles that each will play in getting a job done. That is, that the preliminary discussions, planning, mutual training, etc., are going to be with us for a long, long time. What we apparently have failed to recognize is that the machine itself can be used in such a way that it can help us remember, control, modify, and use the fruits of all this labor much more efficiently.

What should the system programmer be doing? He should not be building systems for himself. He should be providing the standard parts and routines with which systems can be built. Of course, his consultation will be needed, but the whole system literally should be built by the whole spectrum of people who are going to rely on it. How can the programmer anticipate the kind of system and languages that the next project will require? He must be prepared to build any definable system or at least to supervise its building. That is his specialty. The same thing

is true of language. What we need is not a universal language but the ability to define and utilize any language that suits our purposes.

Consider a moment. Systems of 10,000 to 15,000 instructions are routine and 25,000 to 50,000 instructions are not uncommon. Some plans now underway likely will require 50,000 to 75,000 instructions. Where do we stop? Some of these systems are supposed to be all things to all people, but on closer examination it turns out that they generally do about one thing well. Maybe they compile algebraic statements, or maybe they assemble and execute handwritten programs, or maybe they perform certain standard DP functions. But they are only straight-line cuts out into a whole field to be harvested. Furthermore, they are difficult to understand and require a staff of specialists to service each one. Far from facilitating communication, they vastly complicate it.

Suppose then we begin by trying to extract from all systems those parts which are common. We will, without question, have to establish certain rules and conventions; for example, the card form for instructions to a monitor (any monitor) and the first and last words of each tape record, etc. In other words, we have to establish standards. But these standards must not be restrictive to the user. Their only purpose is to allow standardized parts from which any system can be started. We must not forget that knowledgeable people must be integrated with systems if versatility is to be maintained.

In addition to a package of standard parts, something more is needed. This is a standard philosophy of using them. There is a somewhat subtle distinction here. We do not want to make all systems to be just alike, otherwise we may as well program a standard system. The idea is to have standard techniques to cause a system to grow in whatever way is needed at the time. The word "grow" is not used lightly. Certain characteristics of an organism must be provided. To me, the notion of a universal computer-oriented language has meaning only in this dynamic sense.

There is one further requirement. The method must not lead to unnecessarily cumbersome systems, that is, this must not be inherent. With the present state of the art, at least, this will require some rather stringent conventions regarding symbology. Hopefully, they can be relaxed as experience is gained. I do not see this as a great problem, however. People are pretty good at picking up rules of this kind and at getting used to cryptic notations. They may grumble, but their real problems usually lie elsewhere. If they understand the structure of the system and its attendant language, the actual encoding is not too important.

What are some of the features that might be expected to be available with such a method? I think the following is a minimum list:

1. A core of highly polished utility subroutines, a standard skeleton monitor, and a basic assembler; all capable of expansion in well-defined ways and maintained by a single small group (per installation) of highly skilled professional programmers.
2. The capability of defining hierarchial sets of macro-operations tailored to the job and understood to the depth required by the various echelons or categories of professional people involved. Since they would be involved in defining them, they should easily understand them.
3. The ability to make changes at any level, provided that at least the next lower level was understood, and to add levels as need dictates.
4. The ability to debug in detail at object time without detailed preplanning.

5. Complete revision of a system by revising definitions and restarting from the basic system, with a minimum of elapsed time.
6. Partially automatic documentation of a system since higher levels of operations will require fewer lines of programming.
7. Ability to experiment with systems as such, and to extract production systems at will.
8. Alleviation of the training problem by spreading the load now imposed on programming over a wider variety of skills.

Let me re-emphasize that such a technique as just outlined is going to be very demanding of the people using it. Responsibility is a necessary counterpart to responsiveness. The ability to exercise higher level control implies the ability to make bigger, more disastrous errors. With proper safeguards, however, this is somewhat compensated for by the ability quickly to reconstruct a previous situation.

To sum up the foregoing very briefly, no one can become expert in many fields simultaneously. Programming must provide access to machines for others, via the programmer but not by the programmer on every occasion. To expect to find individuals who can understand all the goals, restraints, techniques, and policies of a large project and still be capable of translating these into a mechanized system is unrealistic. To hope that a system can be defined by one group and turned over to another for implementing is just as unrealistic. Everyone must play his proper role from the outset, and communication must exist between members of a team, at the least by pairs. After all, systems exist to serve some human purpose and not to become self-sufficient automatons. The trick for making the hardware a useful member of a team is to construct systems which are self-adaptive in a machine-oriented sense, but externally responsive in a problem-oriented sense. The know-how to accomplish this exists.

Let us turn briefly to another question which might have been taken up first. Are the characteristics of so-called scientific computers necessary in data-handling machines? One might attempt to list the characteristics of both, side-by-side, and compare them. However, this would be tedious and lead to endless debate. I think the question can be short-circuited by the following observations. The only tough problems in scientific computing are the following:

1. What mathematical technique is adequate for and best suited to mechanization?
2. What numerical difficulties can be expected with loss of precision or inaccuracies in approximating functions?
3. How can the data and the problem parameters best be input to the machine, both in meaning and in format? In other words, what do we want to consider as independent variables and how much preliminary transforming and processing of data do we want to automate?
4. How can the flow of data be arranged to minimize running time?
5. How can interruptions, restarts, and output be provided for in order to balance risk against expense?
6. How can work be scheduled to best serve the users within available capacity.

The first two of these are analytical questions, though an understanding of the way in which computers work is implied. The last is a management problem, but relying on machine know-how. The others are all data-handling problems. In short, these questions are equally pertinent to any

large-scale application of computing equipment. Even the question of precision cannot be ignored in data processing, as you will know if you have worked with statistical tables.

In conclusion, then, the question of hardware for logistics research or any other purpose is thrown back at us. Hardware exists which we have not yet learned to use effectively. Until more people take an active part in designing and utilizing actual working set-ups, the question of what more is needed is a vacuous one. Technical developments in information-handling hardware are, of course, continuing. Rest assured that the manufacturers will be responsive to well-defined requirements placed on them. But the primary problem is, as always, the proper organization of human effort to accomplish purposeful tasks with the facilities available. A man I once worked for used to say that a good mechanic never complained of his tools. He was right.

* * *

THE REFLECTION OF LOGISTICS IN ELECTRONIC COMPUTER DESIGN

E. W. Cannon

National Bureau of Standards

Analytical and operational logistics seem, at first glance, to require widely differing computational tools. When one looks more deeply into typical undertakings in logistics, including such varied activities as: fundamental logistics research, the playing of war games, production scheduling, and the determination of supply requirements, appropriate computational techniques do not appear to require markedly differing computational facilities. If differences are deemed essential, they appear to lie in accessibility of storage and the nature and extent of peripheral equipment rather than in the basic organization and mathematical power of the central computer.

Representative problems in varied aspects of logistics are discussed briefly. An appropriate mathematical model is displayed, in each case, with perhaps a heuristic derivation, and is viewed from the standpoint of the computational facilities it requires. New applications of electronic computers, such as the mechanical translation of languages and information retrieval (now in the exploratory stage), which promise to have significant logistical implication, are discussed briefly.

From the standpoint of computational requirements, logistics, both in operation and research, is an extensive field. The pace of developments in warfare, the need for preparedness to act swiftly and effectively lest a "cold" war become hot, the magnitude of the defense effort essential to maintain a state of preparedness deterring potential aggressors from attack, and the dependence of our national economy upon our military budget—these are some of the factors which accentuate the importance of the role of military logistics in the effective use of our resources. The problems relevant to logistics and which must be solved, indeed quickly in emergencies, range over a wide variety of human endeavor and often require sophisticated approaches by way of complex mathematical and statistical models.

The modern electronic digital computer is a useful tool for the investigation and solution of problems which are not amenable to standardized procedures. It is interesting, and relevant, I think, to recall briefly the motivation of the development of this astounding equipment.

We think of the British mathematician, Charles Babbage, as the first to attempt to construct large-scale, automatic computers. In one sense, his environment in the early 19th century was strikingly modern—his projects, first for a differencing engine and later a calculating engine, were supported by grants from the Crown. In another sense, more relevant to his progress, his environment was unfavorable; the engineering art of his time was unequal to the demands of his design, even though the design was functionally feasible.

As a mathematician Babbage was interested in general mathematical computation, not in data-handling or the business problems of his day. The difference engine was intended

to be a special-purpose mathematical instrument, the analytical engine a general-purpose and quite powerful mathematical tool.

We next must turn to the relay computers developed around the middle of the 20th century by the Bell Telephone Laboratories on the one hand, and Harvard University and the IBM Company on the other. These computers were general-purpose, so-called, scientific computers, designed for the calculation of mathematical tables, the solution of differential equations, and the like.

Soon after, the ENIAC, the first electronic digital computer, was developed by the University of Pennsylvania, with the support of the Ordnance Department, U. S. Army. ENIAC was intended to be a special-purpose machine, to be relegated to the computation of firing tables, a specialized differential-equation problem. This computer was placed in operation as recently as 1946—and about 125 years had intervened since Babbage began to work on his automatic "engine."

For the next few years, the impetus to invent or develop fast, large-scale computers was mainly the desire to extend numerical computation to problems called scientific, or engineering, or mathematical depending primarily on which one of the three professions was the speaker's. No one was vocal concerning a need for more powerful computational tools in accounting or other business applications, and the terms data-handling and data-processing had not been coined.

But it was becoming apparent to some progressive managements that high-speed computing might eventually be an effective facility in competitive undertakings; and business, notably life insurance companies, showed interest in the next large-scale electronic computer to appear—the UNIVAC. The UNIVAC design was aimed at census-type applications. It turned out to be very useful for statistical calculations, scientific and engineering problems, and accounting purposes.

UNIVAC was the first large-scale commercial electronic digital computer. It was an electronic realization of the Babbage analytical engine, with refinements possible because of the inherent flexibility of electronic circuitry. It was not necessary to prepare the machine for a problem, as it had been for the ENIAC at great cost in time. It was sufficient to provide instruction and data tapes. And UNIVAC was able to compute simultaneously while operating its input and output auxiliary equipment. This computer proved to be a quite satisfactory calculating tool, having tremendous effect in the development of modern numerical methods.

SEAC, developed by the National Bureau of Standards and intended as a prototype to prove the feasibility of large-scale, flexible electronic computers, predated UNIVAC slightly. SEAC became a productive computer, employed largely on scientific problems but also to considerable extent on business applications, and what were called data-handling problems. This computer was constructed by the Bureau with the support of the Air Comptroller's Office. The motivation of the Air Comptroller was a desire to mechanize and speed up program planning, a logistics-flavored application.

Hard on the heels of SEAC came Whirlwind I at MIT, an extremely fast computer designed to be a part of a flight-training system and to perform "real-time" computing. Then there followed ILLIAC, a University of Illinois computer, modelled on a design created by the Institute of Advanced Study, Princeton. The Institute gave birth to MANIAC a bit later. ILLIAC, MANIAC, and others of this class were designed for scientific application, and the

nearest they came to so-called data-handling and business applications was the reduction of military range data.

A computer having special significance in the perspective of this Conference is the Logistics Computer, which was constructed by Engineering Research Associates with the support of the Office of Naval Research. This computer was designed for logistics research. It has been employed by the Logistics Research Project on a variety of investigations. Mr. Fennell will describe, I expect, the use of this computer by the LRP in his talk tomorrow.

The list of computers of the kind we are contemplating is too long to enumerate fully here. I shall stop with mention of SWAC, built by NBS and now in use at UCLA, the 1100 series of computers of Engineering Research Associates (later part of the Remington-Rand Division of Sperry-Rand) and the 700 series of the International Business Machines Corporation. Among these machines, only the IBM 702 and 705 were considered primarily business machines—the others were more or less oriented to the so-called scientific computation.

Now a striking feature of this list of electronic computers is the fact that, with few exceptions, the computers were designed with the so-called scientific problems in mind—involving both ordinary and partial, differential equations, matrices, the evaluation of multiple integrals, and the like—in disciplines like nuclear and atomic physics, thermodynamics, and mechanics. In these scientific computers, the objectives of an exceedingly fast arithmetic unit and a low-access-time internal memory have tended to overshadow other design features. Significantly, the two exceptions, UNIVAC and the Logistics Research Computer, are a reflection of computational needs in logistics. The Logistics Computer was designed specifically for logistics research, whereas the design of the UNIVAC, was influenced by the requirements of the program-scheduling problem of the Air Force previously mentioned. I was well aware of this facet of the UNIVAC-development picture through serving as a scientific officer on the contract between the National Bureau of Standards and Eckert and Mauchly (who fathered UNIVAC), which culminated in the specifications from which the computer was constructed.

It is of some value in our context to give more attention to the Air Comptroller's scheduling problem. I recall hearing Dr. George Dantzig (then in the Air Comptroller's office) discuss his search for suitable mathematical techniques for the solution of this problem with Dr. John von Neumann at the Princeton Institute for Advanced Study. At the National Bureau of Standards under Dr. Dantzig's direction, we performed numerical experimentation on punched card equipment, which Dr. Dantzig used as a guide in the development of appropriate mathematical techniques. These techniques were to become known as Linear Programming. They were found applicable to important problems, such as: production scheduling, bid-evaluation, transportation scheduling, and even theoretical approaches to war gaming which at first glance appeared to be essentially dissimilar. Moreover, linear programming was proved to be equivalent to certain games, in the sense of von Neumann and Morgenstern, and its techniques have been widely employed in the solution of such games.

A linear-programming mathematical model is easy to describe. It consists of a linear form to be maximized subject to linear constraints of the variables. It is a maximizing problem, simply stated, but not tractable to the methods of the calculus.

An illustrative example might occur in production scheduling. Suppose a company manufactures two items, A_1 and A_2, and wishes to schedule their production to maximize its profit. Suppose further that three machine tools, B, C, and D, are used in the production.

The following data is at hand:

Lot	Profit
A_1	P_1
A_2	P_2

Article	Machine Tool Time Required per Lot		
A_1	B_1	C_1	D_1
A_2	B_2	C_2	D_2

Machine Tool	Time Available
B	T_B
C	T_C
D	T_D

Suppose the unknown number of lots of Article A be designated by X, Article B by Y. Then, considering machine-tool time required, we see immediately that

$$X B_1 + Y B_2 \leq T_B,$$

$$X C_1 + Y C_2 \leq T_C,$$

$$X D_1 + Y D_2 \leq T_D.$$

And the profit, to be maximized, is

$$Z = XP_1 + YP_1.$$

The computational operations involved in the solution of this problem, according to the widely applied simplex method developed by Dr. Dantzig, are quite similar to those pertinent to the solution of systems of linear equations. Scientific computers which can handle what mathematicians call the matrix calculus are therefore suitable tools for the application of linear-programming techniques.

We pass on to a brief consideration of other problems relevant to logistics research. Suppose we list areas of primary concern in logistics, together with kinds of research which have bearing upon them. Such areas are:

1. Requirements
2. Production
3. Supply
4. Communications

Obviously these cannot be considered independent, but close attention to their interrelationship is not essential for our purposes.

In area 1, Requirements, it is essential to investigate the possibility of formalization, for speedy and automatic use in critical situations, of the effect of command decision and operational environment upon logistic requirements. Such an investigation has been carried out by the Logistics Research Project and the Project has evaluated and proposed a three-factor scheme for the application of automatic computers to this problem. The computation involved is straightforward arithmetic. The major emphasis on computer design resulting from widespread use of this method would be effective and fast input-output facilities and ease of inserting the appropriate instructions into the computer.

The close relation of requirements to command decision provides a natural impetus to the study of war games in logistics. The mathematics involved here is quite different from that which we have discussed so far. It may include probability and statistics, differential-difference equations, and the von Neumann-Morgenstern theory of games. In computer design, the premium is placed on arithmetic flexibility and speed of operation. The appropriate computer is a general-purpose, scientific machine of modest input-output capacity.

In area 2, Production, we have already given an example of the kind of mathematical model which is useful (the associated mathematical technique being called Linear Programming).

Area 3, Supply, is obviously a fertile field for the employment of research. To maintain a sufficient supply of material and personnel, on a scale almost beyond comprehension even in peace time, inventories must be kept in order; materials and equipment must be inspected, preserved and overhauled, and transported where needed; and men must be trained and deployed for effective utilization. The associated problems are legion and difficult. Merely keeping account of inventory transactions on a digital computer, or by any other means, by use of Federal Identification numbers is extremely time-consuming because of the fantastic number of items involved. (You will hear later of modifications in the Logistics Computer to increase its power in this respect.)

In the manifold aspects of supply, the value of well-directed research is immeasurable. To mention just one kind of ever-present question (extremely difficult to handle): what is the underlying distribution of life-expectancies of manufactured goods being subjected to acceptance testing? If the acceptance tests are based on a distribution which does not exist, the receiver of accepted goods may be running risks he would, if aware of, take every means at his disposal to avoid.

If, for example, his acceptance testing should be based erroneously upon say, the Poisson instead of the Weibull distribution, he could unwittingly be placing emphasis on extremely long-lived miniature bearings and accepting an unnecessarily high percentage of short-lived bearings—the bearings to be used in the control of the first stages of rockets.

Again, to have a chance to plan optimal inventory policies, i.e., to determine when and how much to reorder at a supply warehouse, one needs to know something about the distribution of demand of the items stocked.

These questions are difficult to handle. A straightforward approach is to determine theoretically what characteristics samples would have for possible underlying distributions and to analyze sufficient samples to reduce the risk of fallacious conclusion to an acceptable limit. Where hundreds of thousands of items are involved, and the sampling analysis must be superimposed upon a heavy daily operational load, this would be a program of awesome magnitude.

Nevertheless the needs to increase the reliability of systems and to develop economical inventory-handling are so great that I believe this problem will eventually be faced squarely.

I predict that it will be handled by the automatic forwarding of data over communication networks by operational digital computers to a central research computer, which will maintain up-to-date analysis of the samples received. A further responsibility, I think, of the central research facility will be to promulgate modifications of testing specifications and inventory policies, when deemed essential on the basis of its analysis.

An important objective in supply activity is the minimization of the cost of material. Attainment of this objective in the procurement of articles of clothing, for example, is difficult to achieve. The cost includes cost of Government-furnished raw material, of its shipment to the manufacturer and of shipment of the finished product to storage points. The picture is complicated by the existence of suppliers of varying output capacity (very seldom can a single bidder furnish more than a portion of the articles needed), and by the practice of differential bidding. According to law, the awards are to be made so as to minimize the total cost to the Government. This problem has been handled on electronic computers according to a linear-programming model. The appropriate computer appears to be a general-purpose machine, with modest input-output capacity.

Transportation itself presents problems of such variety and magnitude that I probably should not have subsumed it under supply. Optimal dispatching from the standpoint of matching supply with need is a tremendous job. The needs do not remain static and neither does transportation capacity in time of conflict. In fact it is necessary to provide, in present times, facility for changing transportation schedules, on a national and even global scale, on very short notice. Requirements are imposed on communications which were unimaginable only a short time ago.

The present world atmosphere makes it necessary to provide information pools on which decisions can be made swiftly in all spheres of military endeavor in case of sudden change in the Nation's defense posture. A feasible method might be the maintenance of information centers of vast storage capacity. However, vast storage could mean intolerable delay in searching files for the information required. For this reason, and because of its high operational speed, application of the electronic digital computer in the retrieval of information is worth exploring. And, indeed, even in our larger libraries and among the professional societies, the problem of information handling has reached a magnitude which has focused attention on the utilization of electronic computers for its solution.

Information-retrieval has many ramifications. How can the information be coded to achieve compactness and freedom from the possibility of misinterpretation when recalled? What kind or kinds of automatic search procedures would tend to minimize search time? Would it be possible for the computer to respond usefully to other than stereotyped queries? And so on !

We do not know an appropriate, inclusive, mathematical model for this automatic handling of information. We are inclined to believe that a closely related problem is that of the automatic translation of languages by electronic computers. For, in one respect, an information retrieval machine must translate an artificial language, its own, into that of the human being desiring information.

In research in mechanical-language translation, some groups are attempting to use a machine in experimental translation—which they hope will deviate more and more from word-for-word substitutions—to derive an appropriate model of the foreign language. Others are attempting, rather, to rewrite known grammatical rules in a form suitable for automatic manipulation by the computers. One of the latter groups appears to have arrived at a promising

model for the syntactical analysis of the Russian language which will probably give faithful machine translation of Russian into English, if no polysemantic Russian words are involved. On the basis of this model the kind of computer required would be a fast, high-speed machine, with a voluminous fixed and reasonably accessible storage for a glossary or dictionary.

Now, we all realize that I have only skimmed the surface in this brief survey of problems arising in Logistics and calling for research and in mentioning appropriate computer attributes. It is highly significant that a logistics problem both influenced the design of an early commercial computer (UNIVAC) and gave rise to a new mathematical discipline (Linear Programming), which has had far-reaching application in numerical computations and theoretical mathematical analysis as well. The logistics-type applications I have mentioned would appear to influence the balance of components of general-purpose electronic computers, rather than call for the incorporation of novel features in computer design.

We have been thinking mainly of research. Now, if we include operations and field activity, I believe that Logistics will add weight to the need for rugged, portable, electronic computers suitable for field operation. In my opinion, both research and day-by-day operation demand concerted effort to determine the basic distributions of inventory demand and equipment-failure times, which, I believe, will eventually result in the establishment of large, central, computational facilities for statistical research.

* * *

AN AUTOMATIC ADDRESSING DEVICE*

J. P. Fennell

Logistics Research Project

In many computations the identification of an item or items requires several decimal digits, an index which cannot be accommodated readily as a machine address. It is characteristic of these indices that the cardinality of the actual range is within the limits determined by conventional high-speed storage. This fact is exploited by means of the system under discussion.

The device[1] to be discussed consists chiefly of a magnetic drum with a capacity of about 400,000 decimal digits with variable word lengths of 4 to 64 digits, a register (FR) with capacity varying from 4 to 64 digits, and an address register (FADR). Upon the execution of a SEARCH command, the entire surface (about 30,000 words at WL=12) is scanned in 32 ms, and, if the contents of FR are found anywhere, this information is made available in the form of an output which can be used to cause a transfer—similarly, if this number is not present. When the number is found its location is entered into FADR. This system may be employed as a conventional storage device with the usual READ, WRITE instructions. The various commands and logical elements of the FDS are discussed.

Quite often in data processing one encounters problems in which a long index (more than 5 decimal digits) must be converted into a computer address. This problem has been met in a number of ways, most of them built around a randomization process based upon the contents of the index. Given a particular job, unicity is achieved in a chain of one, two, or three stages. One has a programming problem and the program which suffices for one situation may have to be overhauled for the next. These problems have another characteristic—the actual range of indices has a cardinality much smaller than does the potential range as determined by the length of the index. Thus, in a job based upon an 11-decimal digit index only, say, 90,000 instead of 10^{11}, actual indexes may occur. The device, which will be described below, solves the problem cited by exploiting this property. The hardware employed is known as the FIIN DRUM System or FDS, a proper but perhaps an unnecessarily restrictive cognomen. The basic idea is that of Professor C. B. Tompkins. The fundamental notion was transformed into basic logical specifications at the Logistics Research Project in 1957-58. These requirements were converted into circuit design and eventually into hardware by the Advanced Electronics Manufacturing Corporation (now a division of Electrada, Incorporated, Los Angeles) in 1958-1959. The system was delivered to the university in December of 1959

The notation in the parent computer as well as in the FDS is excess 3 binary coded decimal. The principal logical elements in the latter are two registers: FR (the FIIN register). FADR (the FIIN address register), and a magnetic drum (FD). FR is of variable length, currently 4 to 18 decimal digits, with provision for extending the upper bound to 64. FD may

*Research was supported by the Office of Naval Research.

[1]Called the FIIN DRUM system.

be written upon or read from in any of these word lengths (WL). WL is manually preset during a problem and, although the results of changing it during a run are known, this practice has been of little interest in applications. FR and FD need not have the same length during computation. FADR holds six decimal digits—10^5 is insignificant for reading and writing. FD holds about 400,000 digits and revolves at 3600 rpm. The packing density is about 225 bits to the inch. Since the drum diameter is 12 inches, this produces a bit-rate of about 500 kc.

The FD has some rather unique physical characteristics. Actually, there are three drum-cylinders bored conically inside. During operation, one of these revolves with a drum spindle while the other two are suspended from an indexing assembly which serves to locate the cylinder being inserted or removed relative to the read-write head banks. Motion is imparted to the cylinder in use by a rubber ring at the spindle top. Air is forced between the cylinder and spindle during rotation. This has the effect of raising the cylinder slightly and producing an extremely even rotation. The operating-head clearance is unknown but is believed to be about 0.00025 inch. When a cylinder is removed or inserted, the head banks are retracted and during operation are located properly by an air-servo system. Each column of heads is controlled by a pair of air jets: one at the top and one at the bottom of the column. The drum system proper was manufactured by Ferranti, Ltd. of Toronto, Canada.

Before discussing the FDS instructions, it may be well to look at the FIIN DRUM addressing system and parallel search feature. The former is shown schematically in Figure 1, the latter in Figure 2.

These instructions are as follows:

READ: The contents of the cell whose address is in FADR are inserted into FR. This is a serial operation so that, if the length of FR is say 6 and the drum WL is 12, the most significant 6 positions of the word appear in FR. In the reverse situation, the word on the drum is inserted into the left-hand side of FR, and the former positions (FR) $10^{11} \ldots 10^6$ are shifted into the right-hand half.

If the address in FADR is illegal, a READ OVERLOAD is generated. This may be used to cause a transfer in the computation or to cause certain other commands to be executed.

WRITE: The contents of FR are stored in the cell whose address is in FADR. The behavior when the WL of FR and the drum do not agree corresponds to the performance in the case of the READ instruction. The overload system is analogous.

SEARCH: The contents of the drum are scanned in two revolutions for the contents of FR. If this number is present on the drum, the address of the first encounter is entered in FADR and a SEARCH NO OVERLOAD OVERLOAD occurs. This overload may be used to cause a transfer as well as certain other operations. If the number is not present a SEARCH OVERLOAD with similar characteristics occurs.

SEARCH FOR 9's: A string of 9's is the normally erased state of each drum cell. The operation is carried out independently of the contents of FR. In all other respects, the operation and overloads are analogous to those in the case of SEARCH.

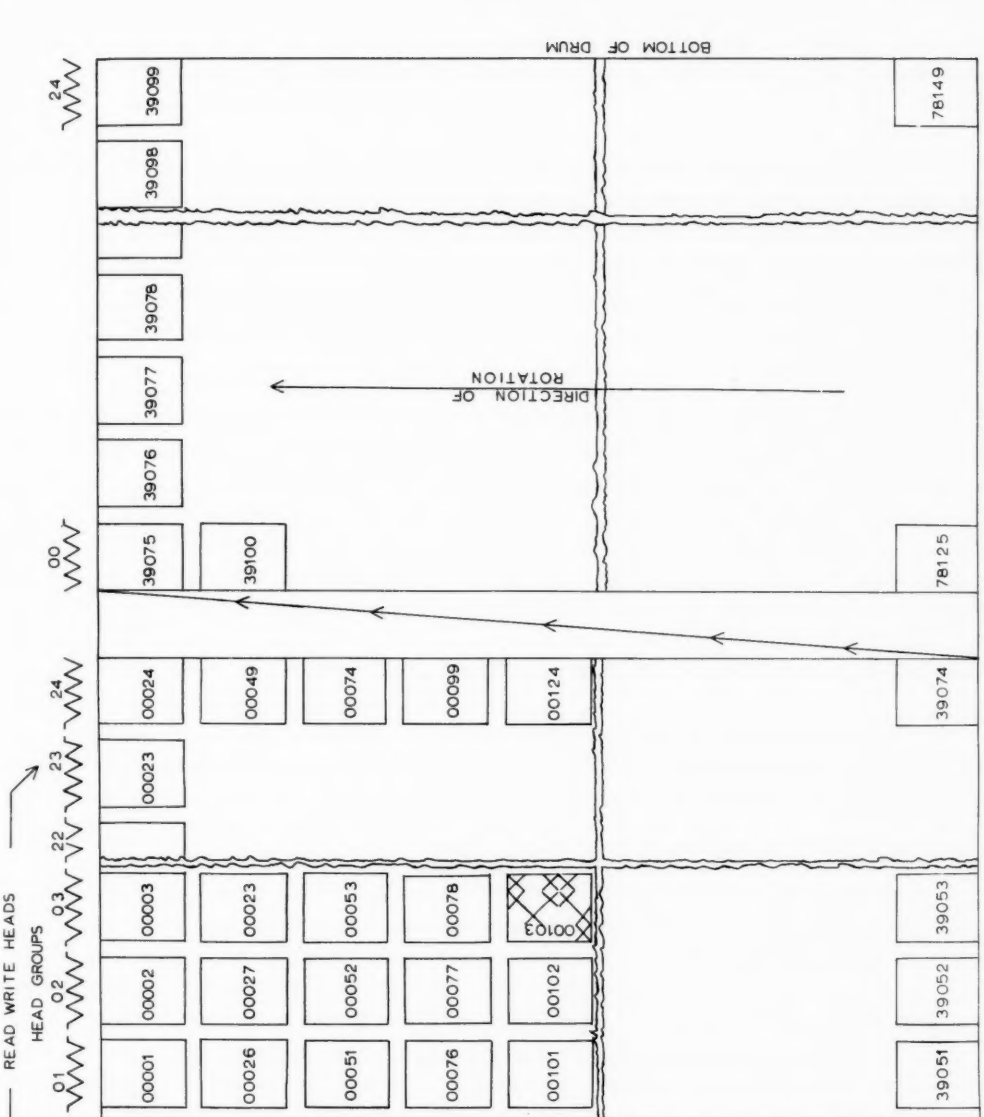


Figure 1 – Schematic of FIIN DRUM addressing system at WL = 5

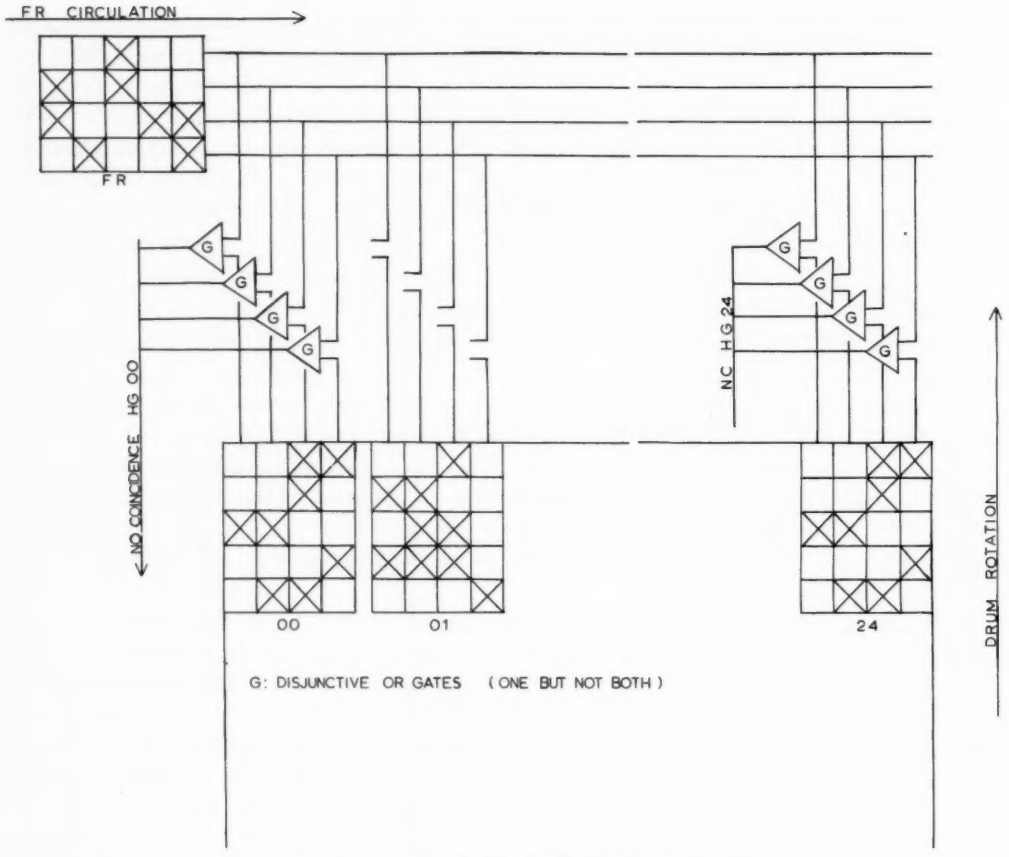


Figure 2 – Schematic of FIIN DRUM search system

Both SEARCH and SEARCH 9's start at the first opportunity, i. e., the first coincidence under the heads is detected. This is in contradistinction to the next two instructions:

SEARCH FROM LAL: It may be desired to obtain the coincidence with the "lowest" address. This command is analogous to those above, except that it is enabled only as the drum comes to reset position.

SEARCH FROM FADR + 1: When an index is stored in more than one place on the FD, it may be required to pick off the addresses for all coincidences pertaining to a given index. The search procedure is started with, say SEARCH FROM LAL, and a coincidence is obtained, say at 01379, and FADR is set equal to 01379. Searching (for the next coincidence) is continued with SEARCH starting at 01380, i. e., all coincidences below this address are ignored.

All search operations are conducted in parallel over half the drum, the dividing line running circularly around the drum center. As shown in B, the 100 lines corresponding to the bit positions under head groups 00 - 24 are opened simultaneously and feed the gates G. All 25 sets of gates are fed simultaneously by the digit being circulated out of FR to the right. At the beginning of each word bank 00 - 24, 25 - 49, 50 - 74, 75 - 99, and each angular location all outputs from the gates G are set to coincidence. By disjunction a particular head group is set to NO COINCIDENCE. When the interstice between two successive word banks is encountered, any head groups not set to NO COINCIDENCE must therefore coincide with the lateral locations of coincidence. The leftmost such location is selected and entered into FADR, i. e., say 13 or 38 or 63 or 88, depending on the bank involved. This is followed by the insertion of angular location and finally by one 0.

In the situation depicted in B, locations 00 and 24 would be left at COINCIDENCE.

Search continues from the side on which it commenced via a switching operation which transfers the scan to the other half during about a 6-bit time reserved for this operation. SEARCH LAL, SEARCH FADR + 1 always start of course on the "left."

Coincidence may be preset on any subset of the positions in the word in FR either manually through a set of CATEGORY SELECTION SWITCHES or under program control by inserting a nonnumeric character in the positions desired in FR. The effect is to suppress search on the positions selected. Thus, one could locate all words beginning with, say 123, irrespective of the composition of the rest of the word.

At WL 10 the drum capacity is of the order of 35,000 words. With a scanning time of 32 ms this amounts to looking at individual words in less than one microsecond.

Some of the more obvious developments which a system of this kind could undergo are: (a) A complete drum scan in one revolution. The present logic is essentially compatible with this purpose. (b) Extension of the size of the system. For example, by means of (a) above and the employment of two drum surfaces, the capacity would be doubled with no loss of speed. (c) Conversion to a core-based system. This would, of course, produce a scan of phenomenal speed.

The exploitation of this system is expected to proceed along at least three lines: (1) as a data-processing aid, (2) as a logical element defining entry points for various programs or subprograms, and (3) as an aid in the construction of logic-type programs—games-playing, "learning," and so forth.

* * *

THE MARRIAGE OF ON-LINE HUMAN DECISION WITH COMPUTER PROGRAMS

Herbert M. Teager

Massachusetts Institute of Technology

As the logistics problems of current interest become more and more complex, it becomes apparent that heuristic methods hold more promise than exhaustive or algorithmic methods of solution. It is also becoming clearer that heuristics, per se, are useless in approaching large combinatorial problems (via pattern recognition, for example), unless some general techniques are developed to draw out the specific "best" heuristics required for each problem. It is a fact that humans are, at present, far more efficient in developing and applying such heuristics than a machine. It is therefore of considerable value to explore how the best features of both human and machine decision can be married. The general problem requires the development of common meaningful languages, specialized input-output display and entry facilities, and a means of time-sharing the computer, in order to efficiently match the machine-vs the human-decision time.

Over the past decade, application of the large-scale digital computer has led to impressive accomplishments in the general fields of science, engineering, and data-processing. Computer application has been most successful in the areas where methods of solution were known, and the dimensionality of the problems were such that it was either unfeasible or impractical for manual methods to be used in carrying out the methods or algorithms necessary for the solutions. Such problems as the solutions of systems of differential equations by numerical-analysis techniques are typical of such application, and in many cases (such as in the computation of flight paths, or in a real-time control system) the slow speed of human calculation was several orders of magnitude below that which was necessary to provide a usable set of answers. This success in applying computers to areas for which they are very well suited has led to the general hope that computers could be applied to areas of intellectual work, other than purely computational and data processing (in areas such as: problem solving, pattern recognition, language translation, theorem proving, and a host of other areas where methods of solution are either nonexistent, hard to find, or impossibly time-consuming to apply). A fairly common feature of these advanced problem types is that they are essentially combinatorial in character; that is, one possible method of solution is to investigate exhaustively all possible alternative choices at each level of decision, and then to select that choice which gives the best result. Unfortunately, the factorial growth of solution-time on such types of problems precludes a straightforward approach, since even the fastest computer presently contemplated would have to spend several lifetimes exploring exhaustively all the possible arrangements of just 20 objects. It is as a result of this "combinatorial barrier" that advanced programming groups have looked with increasing favor upon the development of heuristic programs and programming techniques, whereby a computer could be made to exhibit "common sense"— by following the human technique of eliminating the nonfruitful paths a priori, and only exploring exhaustively the most attractive or potentially fruitful lines of approach.

In some cases, the application of such heuristic methods of solution have led to excellent results, but far more often, finding a suitable heuristic is tantamount to solving the problem in

itself, and a wrong set of heuristics can just as assuredly guarantee that no useful solution will ever be found. Before yielding to despair, let us re-examine the tools at our disposal. Computing systems are, within their own province, highly effective and efficient instruments for certain types of manipulations. They are extremely fast; can have access to vast quantities of exact information; and can perform a large number of logical and arithmetic operations, at a very low cost per operation, but with a high cost per unit of time. The member of the computing system, whom we normally overlook, is the human analyst. He, while no match for either the speed or exactness of his slave, has other compensating features. He is in general lazy, and as a result has no appetite for performing needless calculations, if he can see an analytic or a shortcut method to provide him the needed answer to a problem. He is neither as exact or as literal, and thus is far more likely to correct his own mistakes and misassumptions. He can be extremely good at recognizing abstract patterns, and given a sufficient chance to become acquainted with a problem can many times apply insight to get a "feel" for "what is important," and "what is going on." There are, in addition, a great number of pattern-recognition type of activities that the human is capable of exercising with ease; whereas the same problems are nearly impossible for a computer of the present-day configuration to solve in any reasonable length of time. Finally, his cost-per-hour is much less than his inanimate counterpart. Rather than considering either the computer or the human as tackling a problem, it seems that a combination of the ability of both human and the computer would be far more effective than either would be in tackling alone the more difficult problems that have been mentioned.

There are, of course, many other areas, besides purely heuristic-type programming, where human interaction with a computer in real time can be a tremendously powerful research and development tool. Among such areas are those of system-design, theoretical analysis, and in fact almost any area where conceptual thinking must go hand in hand with analysis and trial. There is another independent argument for real-time usage, even in the areas where work is presently handled on a batch basis. Human thinking can be made more effective if a shorter time elapses between action and consequences; thus, in programming itself, the inevitable human errors are far easier to find and correct if they are made obvious immediately.

Quite obviously, human intelligence has been working cooperatively with machines for as long as digital computers have been in existence. The problem analysis and programming are in general handled almost entirely by human action. The issue here is one of providing a far-closer human coupling with the machine than is present practice, during both the analysis and programming of the problem as well as during the execution of the program or set of programs (that is, on-line, real-time interaction). If we are to consider a man-machine system as a feasible reality, it must operate in real time; that is, there cannot be an appreciable time lag between the posing of a question and its answer. In this sort of cooperative operation, the computer and human would each handle the portion of the problems for which they were best suited, under the control and direction of the human user.

If we are to propose such a system and philosophy, it is well to consider the full implications as applied, not only to computer hardware, per se, but also as applied to programs and programming philosophies, and to communication systems that are used in linking man to machine. For example, it takes human beings a much longer time to come to decisions than it takes a machine and, as a result, a real-time scheme would tend to be highly inefficient of machine time, and thus highly expensive, unless the machine were time-shared, and preferably time-shared among many other real-time users.

Let us first consider the problem of language and communication between man and machine. It is perhaps easiest for people to express themselves through spoken languages. However, drawings and handwriting are normal and common auxiliary media for more exact expression. Likewise, it seems to be easiest for people to understand verbal and graphical messages, and more so if they have the opportunity of asking questions to clarify that which is not understood. At present it appears that verbal inputs are not yet feasible to a computer, although verbal output can be achieved. Handwritten and graphical inputs, however, are becoming more of a reality.

A third means of getting information into a machine is spelling it out via a mechanical device, such as a typewriter keyboard. Typewriting, although nowhere near as flexible a medium for expression as voice or handwriting (because of limited character-sets and a psychological barrier of translating ideas, formulae and data into key strokes on a machine), is available in the form of existing and feasible devices.

To communicate in the reverse direction, i.e., from machine to human, some form of graphical as well as written medium is required. Human intelligence makes good use of eye-sight in spotting the "interesting areas" and "regions worth exploring" from curves and shapes, whereas the same information in tabular form is far harder to assimilate.

Devices that can provide communication facility of the type described do exist, in general. The typewriter or printer are perhaps the most highly developed of these, but unfortunately, as has been mentioned, such devices are not the best medium of expression between human and machine. Display devices (such as oscilloscopes) and their counterparts for introducing information (such as light pens and light cannons) suffer in general from lack of reliability, high cost, and, since they have been afterthoughts for most commercial machines that incorporate them, a very high penalty in computer time for generating and maintaining a display. In addition, such devices have not been engineered for remote usage, and high-priced wide-band communication links are presently required for such usage.

Problems of the type mentioned, namely the questions of providing low-cost, reliable, input-output display devices, are being worked on; but much progress still needs to be made before this highly effective means of communication with a machine is a workable reality for a many-user, on-line, time-shared system.

The problem of input and output language is one that is receiving a large amount of attention from many workers in the computer field. The problem is certainly not unique to on-line usage of a computational system; however, it is perhaps far more crucial in such a system than it is in a normal system, where there is ample time for coding and decoding machine languages into equivalent, recognizable results and processes. In an on-line system, there will be far more need for the user to express himself in the language of his specific field, and conversely for the machine to answer in equivalent terms, rather than for the user to express himself in a general-programming language such as FORTRAN or ALGOL. In a system where elapsed time would be precious, it would be incumbent upon the computer portion of the system to do whatever translation is necessary.

The problem of input-output devices and languages if properly solved, can assure that a computer system will be a powerful assistance for a single user; and if the over-all system is incapable of satisfying the single user, obviously then it cannot satisfy many of them simultaneously.

As has been mentioned, a computer is a very costly device, and it would be totally unrealistic to assume that a single on-line user could be allowed to have full and complete usage of the machine for long periods of time. The wastage of machine time, during the periods where the human must analyze the results and make up his mind as to the next course of action, can be excessively large due to the very slow rate of human decision as compared with that of the machine. In the past, it has been this factor, perhaps more than any other, that has limited real-time usage to emergency situations. Such a restriction, namely, that human decision time must be considered as wastage because the computer is kept idle, need no longer hold with the present generation of computing machines. Machines can be time-shared, and if they are, such idle periods can be used either by other on-line users or by production programs that need not be run in real time.

The principle development in the present generation of computing machines that is a necessary condition for time sharing (although not a sufficient one), is the ability to "interrupt" a program and cause control to change, based on an external demand. A further development is the provision for some type of memory-protection feature that will allow quiescent programs to coexist in core memory with running programs, without danger of inadvertent damage or access.

These unfortunately are but two of many features that must be provided in a machine to make it suitable for real-time usage. Even though it may be possible to allow several programs to coexist in memory at the same time, it is also necessary to have a fairly large random-access store for programs, and results for all of the on-line users, since it would be highly unrealistic to expect that everything would fit into a single core storage. Further, the many input-output devices of the on-line users would need to be connected to the main computer through a suitable exchange and buffer system, and a system of time-measuring devices or alarm clocks would be necessary to keep track of time allocations and bookkeeping in general.

These, then, are the major hardware features for the type of system that has been described, at least insofar as present knowledge and experience can predict.

The analogy between a computation system and a job shop is a very clear one, and this correspondence has been recognized by many workers in the field. However, until the emergence of time-shared and multiprogrammable machines, there was no problem that corresponded exactly to the job-shop scheduling problem, since the "job shop" was run on almost a pseudo production-line basis. It is clear that the more parallel the computer becomes in operation (and time-shared on-line usage is a highly parallel process), the more critical becomes the problem of scheduling "what portion of what problem is done when." It is clear that real-time usage will preclude running any program blindly until completion, since this could mean a long wait for every user other than the one who had control of the machine. It will be necessary to roughly prejudge the running time of each program segment and, based on some strategy, pick that portion of the program which should be run next. Such a strategy, while backed up by a system of clocks to assure uniform usage among users and nonviolation of maximum running-time limits, would have to consider such questions as the cost of bringing in a segment of a program (thus destroying a program already in core), and where the program should be put (since in general all programs would be relocatable), etc. The penalty for a poor strategy would be a slower running and less efficient system. The gain, on the other hand, that could be achieved with a sophisticated and potent strategy would have to be tempered by the cost of running the scheduling program itself, since this is, in effect, "overhead."

Luckily, portions of this problem can be resolved simply. It takes time for human decision and, in addition, takes an appreciable period of time (in terms of a computer memory cycle) for a human to enter a complete request through an input device. Since no major action is necessary until the request is completely formulated, the computation system need not devote its attention to these incoming requests, other than to put them aside, until they are complete.

These, then, are the system considerations for real-time usage. In addition, there are a host of considerations that will have to be considered in writing compilers and other programming systems to be operated in real time. Perhaps the most important consideration is the fact that it will be run in real time, and that as a result error correction can be a more immediate feedback process, and present-day acceptance of re-runs and re-compilations will no longer be required. Real-time usage is not merely a means of getting quicker access to a machine, and if the full implications of this essentially different mode of usage are considered and taken into account, then appreciable gains can be made even in present areas of batch processing (since this can be in effect a new tool, rather than merely a minor modification to an old one).

From this discussion, it should not be inferred that real-time usage and time sharing in general, are still completely in the theoretical stage of development. For the past year, experiments in real-time, time-shared usage have been continuing on an IBM 704, to which we at MIT have added equipment. The experiments to date have been mainly in the area of on-line program writing and debugging, time shared with a production program. The results have been highly encouraging, both with respect to the feasibility of on-line usage, as well as in the area of providing additional insight into the system, equipment, and programming aspects of time sharing. As a byproduct, we have also proved that on-line program check-out can be far more effective than the normal batch type of processing.

In the future, we are planning to provide an on-line facility for key members of the research staff at MIT, with specially equipped input-output devices in their offices that can be tied in to the main computer by ordinary telephone lines. We are also planning to equip several lecture halls with real-time input-output devices, including devices capable of graphical input and output, such as a large blackboard. In the future we plan to conduct experiments in on-line computer usage as a teaching aid for purposes of working out numerical solutions to problems which instructors ordinarily must bypass, due to their nonanalytical character. In the meantime, however, before any of this can become a workable reality, the problems that have been mentioned must be solved. We have made a fairly good start on their solution, but there still remains a long distance to go.

* * *

PANEL 2
NUMERICAL DESCRIPTION OF THE ECONOMY

COMMENTS ON INTERINDUSTRY ECONOMICS*

Alan S. Manne

Cowles Foundation, Yale University

In writing "INTERINDUSTRY ECONOMICS" [1], Hollis Chenery and Paul Clark have furnished the profession with a uniquely useful work. Within the same covers, they have succeeded in providing both a readable textbook for the beginner, and at the same time a systematic treatise for the input-output practitioner. Two chapters are to be singled out for special mention: Chapter 6, "Testing the Validity of Input-Output Assumptions" and Chapter 7, "A survey of Interindustry Research." Together, these chapters provide a long-overdue survey of the literature on empirical implementation of Leontief's model of the economy. Much of this literature has appeared only in mimeographed form, and has not hitherto been generally accessible.

So much for the nice things that deserve to be said about this volume, and now for my chief complaint: In comparing Leontief's original input-output formulation of interindustry economics with a linear programming approach to the same class of substantive problems, the authors lean over backwards to find merits in the "square" tableau as compared with the "rectangular" one. For example (from p. 299):

> The main obstacle to the ready adoption of linear programming for national planning is the heavy demand that it makes for data. The construction of an input-output table of 40-50 sectors, with a corresponding breakdown of final demand, is a strain on the statistical resources of most countries, and the adoption of activity analysis in a very detailed form is not a feasible task at the present time.

What the authors seem to be saying here—and in other passages as well (e.g., pp. 82, 128, 129)—goes something like this: Undoubtedly the rectangular matrix form of linear programming provides a certain type of theoretical flexibility not inherent in the square Leontief model. (Alternative processes of production are ruled out in the latter model and admitted in the former.) However, it is more expensive to perform the numerical analysis in the case of linear programming, and it is also more expensive to collect the data.

In other words, Chenery and Clark have scrupulously refrained from claiming that the square matrix is defensible on the basis of its theoretical or its predictive merits. Instead, they rely upon the very pragmatic argument of apparent cost. They have also refrained from pointing out that some of the best thinking of input-output practitioners has been devoted to

* Research undertaken by the Cowles Commission for Research in Economics under Task NR 047-006 with the Office of Naval Research.

patching and tinkering with the basic framework—improvisations which could have been largely avoided through activity analysis. Because of the constant need for this patching and tinkering, it is to be doubted seriously whether the activity analysis formulation is really more expensive from the viewpoint of data collection. Indeed—at any given level of commodity detail—it is likely that just the reverse is true. Now to cite some evidence for these rather dogmatic assertions—evidence which is to be found directly within the covers of the Chenery-Clark book:

1. Some of the most interesting applications of interindustry analysis to problems of economic development consist of the problem of choice between importing a given commodity and producing it at home. (Chapters 10 and 11 contain a fascinating series of such "make-or-buy" problems based upon the authors' own experience in Italy.) When the authors attempted to apply the square tableau formulation to a 1951-56 projection of the entire Italian economy, they found it necessary to engage in a substantial "auxiliary analysis" of imports—in effect, trying to outguess the solution of their model.

> For each of the 200 product classes in the table, a marginal import parameter ... was esimtated. Some of these import parameters were in fact based on the average import proportions of the base year, others on a more recent year, still others on marginal trends in imports from year to year, but in principle each was an independent estimate of import substitution possibilities for a particular product class. (p. 254)

By contrast, when Chenery and Clark turned to the use of a programming model for their pilot study of Southern Italy, they found it possible to handle the problem of imports by means of the very straight forward device of stipulating world prices, and then permitting the model itself to adjust the quantity of imports in the light of relative scarcities throughout the system (pp. 290-298). Which of these two approaches necessitated a greater amount of preliminary data processing? It is fairly clear that the advantage lies with the activity analysis formulation.

2. In the 1951-56 projection for the Italian economy, it was necessary to make an "auxiliary analysis" of the iron and steel industry. "... as a result of blast furnace expansion under the Marshall Plan, the steel industry was using more pig iron relative to scrap, and appropriate adjustments were made in these coefficients, as well as in the related input coefficients for iron ore and pyrite ashes" (p. 253).

Would it not have been more straightforward and less time-consuming to stipulate several alternative processes for steel production—some entailing a large amount of pig iron inputs and others a small amount—and then permit the model itself to choose among these, on the basis of relative scarcities?

3. In this same Italian projection, it was found desirable to make a detailed substudy for the major fuels.

> Here it was clear that a significant technological change was in progress, centered around widespread substitution of domestic natural gas for imported coal. The situation was complicated further by interrelations with fuel oil and domestic low-grade coal, ... as well as by differences in the ability of different consuming industries to make fuel substitutions. (p. 253)

This example of fuels substitution should serve as additional evidence in favor of using a rectangular matrix as the basic analytical format in interindustry studies.

4. The problem of by-products constitutes an unmitigated nuisance within a square tableau.

> For example, demand for hides by the leather industry may be portrayed as stimulating the livestock industry to increase its production of meat, rather than the other way around. A much better way to handle a by-product is to treat it as a fictitious intra-industry sale (e.g., hides to the livestock industry) and to calculate a regular input coefficient related to the output of the industry itself. (pp. 145, 146)

That this device will occasionally work may be seen in the following numerical example (positive outputs and negative inputs).

TABLE 1

	Before Fictitious Sale		After Fictitious Sale	
	Domestic Production of Livestock	Import of Hides	Domestic Production of Meat	Import of Hides
Hides	+1	+1.	0	+1
Meat	+1	0	+1	0
Foreign Exchange	-1.	-0.80	-0.20	-0.80

If, however, the foreign exchange coefficient for importing hides changes from -0.80 to -1.20, this method breaks down. The domestic production of meat will then be accompanied by a positive by-product of +0.20 worth of foreign exchange! Again, this a "conceptual" difficulty that could have been sidestepped by the use of an activity analysis format.

To conclude: Activity analysis does little to solve some of the major problems of interindustry econometrics—specifically, the projecting of final demand, the forcasting of technological advances, and the estimating of capacity in labor-paced processes. However, activity analysis does provide a convenient framework for handling certain kinds of difficulties—import substitution, processing substitution, labor-capital substitution, and the output of by-products—problemsthat have arisen in any of the empirical applications proposed by Chenery and Clark. On balance, their book provides impressive evidence against the presumption that the collection of data for a square matrix is cheaper than for a rectangular one.

As for the relative costs of computing, it is a fair bet that advances in the art of electronic calculations have already rendered this an irrelevant issue. (See the Appendix for an estimate of these relative costs.) The "decomposition" principle [2] has opened up the possibility of dealing with models containing literally thousands of equations and of activities. With such tools at our command, it seems a rather dubious use of expensive labor to insist upon fitting interindustry studies into square matrices. The most likely course of evolution will be a progressive increase in the number of sectors analyzed by means of several alternative activities and a corresponding decrease in the number of sectors for which our models contain just a single producing activity. It would be a serious error to underestimate the speed at which this evolution is taking place.

APPENDIX

RELATIVE COSTS OF COMPUTATION

To my knowledge, there is no experimental evidence currently available on the relative computational effort needed in order to solve a Leontief matrix and a linear programming problem. (The 200-sector matrix for 1947 was inverted on the Univac I, but very little experience with general-purpose linear programming on this machine has ever been reported.) The following theoretical comparison ignores many factors quite significant in actual numerical analysis, but should at least serve as a starting point.

In this comparison, it will be assumed that both the Leontief matrix and the linear programming matrix have the same number of rows, m. The number of columns in the former case will also be m, and in the latter case will be denoted by n ($n > m$). To facilitate this comparison, we shall make one highly unrealistic assumption—that all coefficient entries in the original matrices are nonzero. (Sparseness in these matrices will obviously lower the absolute costs of computing, but should not grossly distort the relative costs.)

In the case of linear programming, we shall suppose that the numerical analysis is performed by means of the revised simplex method, and that the basis-inverse is continually maintained in its explicit form. Since a linear programming solution remains feasible only within the immediate neighborhood of a particular vector of input availabilities and of "final demands," it can be argued that such a computation is analogous to a specific solution of a Leontief matrix. However, since an inverse is obtained by the simplex method and since this inverse can be employed for purposes of parametric programming, it can also be argued that the inversion of a Leontief matrix—rather than a specific solution—corresponds more closely to a linear programming optimization. This second viewpoint will be the one adopted here. That is, we shall compare the work involved in a linear programming solution with that involved in inverting a Leontief matrix.

The comparison will be conducted solely in terms of the number of multiplications required at each of the major steps in the process. The number of multiplications is generally regarded as a good indicator of the total volume of arithmetic and of logical operations.

Under these ground-rules, we utilize Evans' estimates for input-output calculations [3]. According to Evans—and contrary to popular impressions— it is more efficient to invert a Leontief matrix by means of the old-fashioned "elimination method" rather than "expansion-in-powers." See also [4]. By use of elimination methods, the required number of multiplications can be kept down to only m^3, rather than something of the order of the $5m^3$ involved in the power series technique. (For a step-by-step explanation of the elimination method, see Reference [5].)

Turning now to the linear programming algorithm, one major difficulty is that there is no easy way to predict the exact number of basis changes needed in order to reach an optimal solution. For this reason, it is convenient to regard the number of these iterations as a parameter r, and to begin by calculating the number of multiplications required at each iteration.

Now at each such step, it will be necessary to make $3m^2$ high-precision multiplications: one set of m^2 multiplications to determine the implicit prices, one set to obtain the "representation" of the new activity in terms of the previous basis, and one set to transform the previous basis-inverse into the new inverse. For more details on this estimate, see [6] and also [7]. In addition to these high-precision multiplications, there will be one group of low-precision operations needed in order to determine which of the n possible activities ought to be the new one introduced into the basis. Since these $m \cdot n$ multiplications may be performed

with low-precision arithmetic, they are likely to be the equivalent of only $(m \cdot n)/4$ high-precision operations. The total number of equivalent high-precision multiplications for a linear programming solution then works out to be:

$$\text{(number of iterations)} \quad \times \quad \begin{pmatrix}\text{number of multiplications} \\ \text{per iteration}\end{pmatrix}$$

$$(r) \quad \times \quad \left(3m^2 + \frac{m \cdot n}{4}\right)$$

Furthermore, RAND's experience suggests that $r \approx 2m$, and that $n \approx 4m$. Inserting these very rough numbers, the total works out as follows:

$$(r)\left(3m^2 + \frac{m \cdot n}{4}\right) = 8m^3 .$$

Summing up: This comparison suggests that the amount of computing effort needed in order to optimize a linear programming problem lies around eight times that required for the inversion of a Leontief matrix. It is noteworthy that the relative costs are unaffected by the absolute size of the matrices involved. It is also noteworthy that advances in the art of computing would today make it possible to invert in less than 15 minutes the 200-order Leontief matrix of the U. S. —an operation requiring 48 hours in 1952.

REFERENCES

[1] Chenery, H. B., and Clark, P. G., Interindustry Economics, John Wiley and Sons, New York, 1959.

[2] Dantzig, G. B., and Wolfe, P., "A Decomposition Principle for Linear Programs," Operations Research, Jan. - Feb. 1960.

[3] Evans, W. D., "Input-Output Computations" in The Structural Interdependence of the Economy, T. Barna (Ed.), John Wiley and Sons, New York, 1954 (especially pp. 77-81).

[4] _____, "Letter to the Editor," Econometrica, Jan. 1958.

[5] Gass, S. I., Linear Programming, McGraw-Hill Co., New York, 1958, pp. 27-31.

[6] Wagner, H. M., "A Comparison of the Original and Revised Simplex Methods," Operations Research, June 1957.

[7] Orchard-Hays, W., "Evolution of Linear Programming Computing Techniques," Management Science, Jan. 1958.

* * *

REPLY TO ALAN MANNE'S "COMMENTS ON INTERINDUSTRY ECONOMICS"

Hollis B. Chenery
Stanford University

I sympathize with Manne's criticism of the continued use of the crude input-output system for empirical work and with his impatience to adopt the theoretically more satisfying conceptual framework of linear programming. He correctly states that "some of the best thinking of input-output practitioners has been devoted to patching and tinkering with the basic framework." However, I cannot agree with Manne's conclusion that these improvisations "could have been largely avoided through activity analysis" or with his implication that the "optimum" solution is what is wanted in all cases.

There are two main reasons for treating choices among sources of supply through "auxiliary analyses" rather than as part of a programming solution: (1) the nature and accuracy of the data, and (2) the limited validity of the optimizing assumption.[1] If it is assumed that the demand and supply functions are known with certainty and that the least-cost solution is desired, as in the pilot model of Chapter 11, there is no question that a nonlinear programming formulation is more satisfactory than an input-output approximation. It should be noted, however, that a linear-programming model is a gross over-simplification of import choices, even if some 200 commodity groups are distinguished, because there is considerable heterogeneity at this level of aggregation. The solution to a linear model will indicate that each commodity group should either be produced or imported in its entirety—a most unlikely outcome in practice—unless additional restrictions are introduced. The same can be said of alternative techniques of production.

If, on the other hand, it is desired to project the actual course of the development of an economy over a five-year period, as in Chapter 10, it requires great faith in the perfection of the market mechanism to assume that the economy will approximate the optimum solution to a programming model at all closely.[2] If data on alternative production activities and imports were available in sufficient detail, it might be feasible to construct a rectangular activity matrix and to predict the rate of adoption of more efficient techniques in each industry. Typically, however, studies of technological change are made only for certain inputs—fuels metals, etc.—rather than for whole activities. This leads to the "auxiliary analysis" procedure of finding a solution by changing input coefficients in the light of technological trends and of the demands estimated in a trial solution.

My position on the choice of models for economy-wide analysis can be summed up as follows:

1. Activity analysis is a very valuable guide to formulating the model and allocating research time among the thousands of relations which need to be specified. A pilot model in

[1] I do not include the cost of computation as a serious obstacle. My doubts about computational feasability apply only to detailed process analysis (p. 128) and not to choices at the plant or industry level. The quotation cited from p. 299 does not refer to the cost of numerical analysis, as Manne suggests.

[2] It would seem desirable to test out the predictive ability of a programming model on single industries, as in Henderson's study of the coal industry, before applying it to the whole economy, where the predicted reaction of one sector depends on optimum reactions in other sectors.

programming form may be quite useful in indicating the sensitivity of solutions to changes in various parameters and thus in guiding further empirical research.

2. The additional assumptions and restrictions that have to be introduced to produce realistic results from a programming model are likely to be more time-consuming than the ad hoc methods currently used. A programming solution will only improve the results when: (a) the objective is policy-making rather than prediction, (b) the data available warrant the more precise formulation, and (c) the choice of activities does in fact depend on the solution to the model.[3]

3. The main empirical limitation to the present use of programming models for the whole economy is the rudimentary knowledge of capital inputs now available. Until these can be specified with some accuracy for alternative techniques, the added precision given by a programming solution is likely to be spurious.

4. The demonstrated advantages of programming models apply to policy determination but not as yet to prediction. Just as regression analysis based on large aggregates has some advantages over input-output analysis in predicting changes in output because of the inclusion of time trends, so will the statistical estimation of input-output and import coefficients from time series offer some advantages over activity analysis in taking account of factors which cannot be analyzed in detail.

These conclusions argue for caution in the use of optimizing solutions for other than illustrative purposes but not against the collection of data on alternative activities in the form of a rectangular matrix. The practical solution to the model is likely to contain a mixture of activities for each output, however, rather than the use of the most efficient one. The solution procedure should thus include elements of judgment which are not readily formalized for an electronic computer.

[3] See the comment on p. 299 on the criterion for adding alternative activities.

* * *

CHANGING OUTPUT-INPUT RELATIONS WITHIN THE NATIONAL ECONOMIC ACCOUNTS

John W. Kendrick

The George Washington University and
The National Bureau of Economic Research

The national economic accounts provide the framework for a comprehensive numerical description of the economy. The most significant work of this decade in U. S. national accounting will be the integration of the related informational systems that have been developed in the past decade into the basic income and product accounts. This paper discusses one of the related bodies of estimates—real factor cost (input) and product (net output) by industry—developed by the author. These estimates are a link between the national production account and the interindustry matrix.

There is wide dispersion among industries and variability over a time period, in the rates of change of output in relation to the factor inputs, singly and in combination. But in almost all industries and over most decadal periods in this century, total factor productivity has risen. Capital has grown faster than labor input, but savings in capital as well as in labor have been recorded in most industries. There also have been substantial savings in purchased materials in many sectors, but in agriculture, purchased products have risen significantly faster than gross output. There are interesting systematic relationships among the several major "partial" productivity ratios.

Logistics and general economic or business forecasting and planning must allow for changing input-output relations, especially over longer time periods. Productivity time-series are a necessary background for projections, but they must be supplemented by special studies.

The most comprehensive numerical description of the economy is furnished by the national economic accounts, broadly conceived to include the various related systems that have been developed in recent years. Many of the possible elaborations—in the direction of sector flow-of-funds, balance sheet, and wealth statements; balance of payments, and intercountry purchase and sale matrixes; industry income and product in constant prices, and interindustry input-output matrixes—have not yet been incorporated into the structure of the U. S. national accounts, but they can be. [1] Much of the significant work in national economic accounting in the decade we are now entering will undoubtedly be devoted to integrating these systems into the basic income and product framework. In the process of integration, the structure of the basic accounts as now conceived, as well as the related informational systems, will be modified and refined. And it is to be hoped that the basic data on which the estimates rest will be extended and improved.

---

[1] See The National Economic Accounts of the United States. Hearings before the Subcommittee in Economic Statistics of the Joint Economic Committee, Congress of the United States (U.S. Government Printing Office, 1957), especially the Appendix containing the Report of the National Accounts Review Committee, National Bureau of Economic Research.

In a broad sense, the national economic accounts are a framework within which significant economic variables can be presented in a systematic fashion, ordered to show their interrelationships in given time periods and through time. Although a frame, the accounts do require an element of value judgment, as in deciding: (a) what activities are economic and to be included, as opposed to the noneconomic activities to be excluded; (b) what goods and services are "final," as opposed to those that are "intermediate" and consumed in the production process; and (c) what are the boundaries of the national or geographical area to be described. Further judgment is required with regard to the economic sectors that are significant in the sense that the units included make their decisions in terms of like stimuli and objectives. It is also necessary to decide what degree of detail is significant and useful.

These decisions are based on judgments as to the ideal shape of the accounts, but they are also influenced inevitably by the form in which data are collected, the margins of error attaching to components, and the costs of collecting data and making estimates in alternative classifications. Nonetheless, we must work towards obtaining estimates in the most useful form for analytical purposes, and not permit the data to determine our information systems in the long run, although they inevitably tend to do so in the short run.

It is my impression that the basic data are, with occasional setbacks, being extended and improved. I venture to guess that they will continue to be extended and improved as the national accounts are elaborated further, reflecting the need for better and more comprehensive economic intelligence as our economy grows in size and complexity. The estimates make possible better analyses of economic and technical relationships, and thus improve the ability of officials both in public agencies and private industry to forecast, to plan adaptations to expected events, or to influence coming events.

In this paper, I should like to present in summary fashion some findings with respect to industry productivity-changes in the U. S. economy. They are based on my estimates for the National Bureau of Economic Research of output, factor inputs, and inputs of intermediate goods and services purchased from other industries (which I shall hencefore call "materials," for short). Any type of economic forecasting or planning, whether for business, logistics, or general public-policy purposes, must take into account changing input-output relations—and the longer the time horizon involved, the more important the productivity projection. Knowledge of past changes is necessary, though not sufficient to the projection. Before detailing the nature of past changes, however, I should like to indicate how the productivity estimates fit into the national economic accounting framework.

THE PRODUCTIVITY RELATION

Productivity may be defined as a relationship—frequently expressed in ratio form—between output and associated inputs in real terms. It is apparent that the current value national product, "deflated" to eliminate the effects of price change, provides a comprehensive measure of the physical volume of production of final goods and services for the economy as a whole. What is not as immediately apparent is that the national income, which represents the money costs of units of labor and capital services, may also be deflated by appropriate factor price indexes in order to isolate changes in the physical volume of factor inputs, or "real factor costs." This has been suggested in U. N. documents,[2] but is not attempted by the Commerce

[2] See particularly A System of Price and Quantity Indexes for National Accounts, United Nations (E/CN. 3/L. 46), 1957, mimeographed.

Department. In my estimates for the National Bureau of Economic Research, I have found it more convenient to estimate the factor inputs directly, and to obtain the implicit factor price deflator as the quotient of national income and real factor cost.

Since national income is defined and measured to equal the national product at factor cost, the real outputs and inputs are consistent and equal in the base period if the final goods and services are accorded unit factor cost weights. The constant dollar input series may be divided into the real product series to obtain productivity ratios through time. In this paper, I shall not go further into problems of defining and measuring output and the inputs in order that we may have time for more interesting matters.

The advantage of relating real product to the sum of associated human and material factor inputs is that the ratios indicate the net saving of inputs—and thus the change in productive efficiency. Ratios of real product to single classes of input, such as labor, reflect the effect of factor substitutions as well. Further, the total factor productivity measures are consistent with the implicit product and factor price indexes. The product price index can be decomposed into unit factor costs of the various types, which, in turn, can be expressed as quotients of the corresponding factor price and partial productivity measures.

The national accounts framework can also be used to obtain industry estimates of real product, factor input, and productivity on a basis consistent with the national estimates. The industry approach is, in fact, a third method of obtaining national aggregates and represents a decomposition of the production account by industrial sector. Real product in the various industries may be estimated as the difference between the deflated values of output and the deflated values of intermediate product purchases—the so-called "double-deflation" method. [3] The interindustry or "input-output" matrix represents an elaboration of the industry production account and is helpful in establishing weights for the price-deflators of purchased goods—although a product breakdown for both sales and purchases is needed to obtain ideal deflators.

In current values, industry product at factor cost, or net value added, is equal to the national income originating. Deflated national income, or real factor cost, originating by industry represents a direct decomposition of the national aggregate and is consistent with the industry real product estimates. Industry productivity estimates, obtained as quotients of real product and real factor cost by industry, are thus consistent with the national productivity estimates. The added industry unit value and the factor price indexes are likewise consistent with the national measures and with one another.

[3] The Department of Commerce has made this type of estimate only for the farm sector. The Bureau of Labor Statistics has estimated real value added in manufacturing (Report No. 100). The basic categories needed for this type of estimate are as follows:

Line No.	Industry A
1	Value of gross output
2	Price deflator
3	Constant dollars (1 ÷ 2)
4	Purchased intermediate products (consumed)
5	Price deflator
6	Constant dollars (4 ÷ 5)
7	Industry products (value added)
8	Current dollars (1-4)
9	Constant dollars (3-6)

NATIONAL TRENDS[4]

Between 1919 and 1957, real private product per manhour increased at an average annual rate of 2.6 percent. If manhours by industry are weighted by average hourly earnings in the base periods, output per unit of weighted labor input increased at an average annual rate of 2.3 percent. The 0.3 percentage-point difference reflects the effect of relative shifts of workers from lower-paying to higher-paying industries.

Real product in relation to the real net capital stock increased at an average annual rate of 1.3 percent over the same period. It is thus apparent that the real capital stock per unit of labor-input increased at an average rate of one percent a year. Nevertheless, savings in capital as well as in labor were achieved.

When the ratios of real product to each of the two major input classes are combined by the income shares of each, to show the net savings of factor inputs, the resulting total factor productivity measure shows an average increase of 2.1 percent a year.

By looking at rates of change between cycle peaks, it is seen that there has not been much variability in rates of change in total factor productivity—the average deviation from the trend rate since 1919 has been 0.3 percentage point. There is more variability in each of the other two ratios. This implies that deviations from trend in output per unit of labor input and in output per unit of capital input tend to be negatively correlated. In other words, in subperiods in which capital per manhour increases more than the trend rate, so does real product per manhour.

For example, in the postwar period from 1948-1957, the output per unit of labor input shows an average annual rate of increase of 3.1 percent compared with the trend rate of 2.3 percent. But the output-capital ratio showed no increase, reflecting an accelerated rate of increase in real capital stock per unit of labor input. Total factor productivity, increasing at an average rate of 2.3 percent a year, thus shows little evidence of acceleration when compared with the 38-year trend rate.

The real national product does not separately include materials consumed in the production process, since their value is already reflected in the value of final products. Also, the further inclusion of intermediate products would involve double counting. Nevertheless, any savings in materials usage per unit of final products over a period of time would result in a faster increase in real product and productivity than would occur if the ratio of materials input to real product held constant. During the period from 1900-1952, the use of raw materials (foods, energy materials, and physical structure materials) per unit of real product was approximately cut in half. This has been due, not just to a more efficient use of raw materials, but also to a higher average degree of processing of materials, and a growing relative importance of services in the national product. I have calculated that declining unit materials consumption has accounted for 0.25 percentage point or about 7 percent of the average annual rate of increase in real national product.

INDUSTRY PRODUCTIVITY CHANGES

For 33 industry groups we have estimates of output, manhours, net capital assets, and materials consumption for the period from 1899-1953, and selected years of high-level activity in between, approximately a decade apart, which we call subperiods. Output-per-manhour estimates alone are available for a much larger number of industries.

[4]The findings summarized in the rest of this paper are based on estimates contained in my study for the National Bureau of Economic Research, Productivity Trends in the United States, in press. For an earlier summary, see Solomon Fabricant, "Basic Facts on Productivity Change." Occasional Paper 63. National Bur. Econ. Research, 1959.

Output Per Manhour

Starting with the output-per-manhour measures, we see a considerable degree of dispersion of trend rates of change over the 54-year period. Average annual rates of increase range from a low of 0.7 percent for anthracite coal to a high of 6.2 percent for electric utilities. Within manufacturing, the range is from 1.2 for lumber and products to 5.1 percent for tobacco products.

The majority of groups have secular rates of advance that cluster about a weighted mean, falling in the 1 to 3 percent a year class interval. No group has shown a secular decline, and the distribution of rates of change is somewhat skewed to the right.

Dispersion of group rates of change in output per manhour is approximately twice as great in the subperiods, on the average, as over the long period, based on a comparison of average deviations. A few declines occurred in each subperiod, although no single group declined in more than two of the six subperiods.

Consistent with these facts, there has been considerable variability in group rates of change in output per manhour over the six periods. The patterns of change vary widely, but acceleration and deceleration of group rates tend to be offsetting in their effect on the economy rate. There has been a tendency for industries with low trend rates of advance to improve their relative standing. In general, the degree of dispersion of rates of change has decreased, indicating a faster diffusion of innovation throughout the economy. For example, agriculture, formerly a relatively poor performer, has shown one of the highest rates of productivity-advance since 1937. The noncommodity area of trade, services, and finance has also improved its performance somewhat since the late 1930's, but is still below average.

Output Per Unit of Capital Input

Despite the substitution of capital for labor in most areas of the economy since 1899, substantial savings in capital per unit of output were realized in most segments and groups. Secular rates of change in output per unit-of-capital input (the reciprocal of the "capital coefficient") ranged from a small negative in lumber and products to 4.7 percent a year in the electric utilities. Half of the group rates of change fell between 1 and 2 percent a year, averaging 1.2 percent.

In most of the segments and groups, the output-capital ratio declined during the early phase of mechanization or building of plant in anticipation of expanding output. The low point has come at widely different times in the industries surveyed. But in all of them, innovations of a capital-saving nature eventually resulted in raising output-capital ratios, with a net increase over the period in most. Little advance was experienced in the period since 1948, however, as a high rate of investment following the wartime restrictions built back capital stock about as rapidly as output was expanded.

Variability in output per unit-of-capital input over the subperiods is 50 percent greater than variability in output per unit-of-labor input. The dispersion of group rates of change in the subperiods averaged 3 times greater than the dispersion in rates of change in the group output-labor ratios.

Total Factor Productivity

Weighted averages of the ratios of output to each of the major factor inputs in all of the 33 groups gives us the pattern of change in total factor productivity, i.e., the relative net savings in factor inputs by industry. Since labor gets the predominant weight in most of the

groups, the patterns are very similar to those described for output per manhour, but with some differences.

In almost all groups, capital per manhour increased over the period, and total factor productivity increased by about 0.3 percentage point less than output per unit of labor input. Most groups fell within a range of 0.2 to 0.4 percentage point lower rates of increase by using the total measure—but in half a dozen industries or so, rates of increase in capital per manhour were much higher, and thus the increases in total factor productivity were considerably lower than in output per manhour.

In general, the degree of dispersion of industry rates-of-productivity change was no greater than that of output per manhour, and was much less than that of output per unit of capital. This indicates a mild positive correlation between rates of change in capital per manhour and output per manhour. Likewise, the variability in rates of change of total factor productivity over the subperiods was a bit less than for output per manhour, and much less than for the output-capital ratio.

Variability in the annual changes in the productivity ratios are substantially greater than subperiod variability. Annual percent changes are substantially higher in periods of expansion than in periods of contraction in the groups for which annual estimates are available, with the exception of agriculture.

The greater stability in rates of change of the total factor productivity suggests that this is a better vehicle for economic projections than the more variable partial productivity ratios. Projections of labor and capital unit requirements can be derived from the total factor productivity and real product projections by explicit treatment of factor substitutions.

Unit Materials Requirements

Historical series on real purchased materials consumption cannot be made with precision, but I have been able to assemble rough estimates for the 33 industry groups. These estimates confirm the impression gained at the national level that output per unit-of-materials input has generally risen. By using the manufactured-foods industry as an example, since the estimates are relatively good, we find that the unit-materials requirement fell by more than one-fifth between 1899 and 1953. Much of this was due to a shift in the product mix of the group towards more highly processed foods, but there was also some materials savings within individual 4-digit industries.

The farm segment stands in contrast to most of the manufacturing groups, however, in that purchased materials per unit of output have risen by more than 150 percent over the 54-year period. As a result, real product (value added) originating has gone up significantly less than gross output.

Correlation analysis indicates that those industries that have greater than average increases in total factor productivity tend to have greater than average savings in materials. This suggests that management efficiency in use of the factors may carry over with respect to efficiency in the use of intermediate products, and/or that the industries with relative increases in productivity also tend to be the ones in which there are relative increases in the degree of processing of materials.

ASSESSING THE PRODUCTIVITY OUTLOOK

I should like to conclude with some observations about the problem of projecting productivity change, by industry, into the near- and intermediate-term future. This is essential

for general economic forecasting, due to the impact of technological change (as reflected in productivity ratios) on economic aggregates and structure. Furthermore, it is necessary for business planning, since industry and company sales are generally related to aggregate levels of demand and activity, and since company resource requirements are related to company production and productivity projections. Productivity projections are also essential for logistics feasibility tests if the time horizon is explicitly taken into account.

Obtaining estimates of past productivity movements is certainly a first step in preparing projections, but it is not enough merely to project past trends—although this would generally be better than making no allowance for changing output-input relations. It is true that at the national level, the coefficient of variation in rates of total productivity increase in subperiods relative to secular trend has been small; but it is much larger for two-digit industry groups and larger still for 4-digit industries. The problem is worse with respect to specific input-output ratios because of irregular changes in combinations (or substitutions) of inputs. These irregular changes result from changing technology, changing relative prices, and other supply conditions. In other words, even if it is correct to assume that productivity advance continues nationally at a moderately steady pace, the industry impact is almost certain to change compared with the longer-term or recent past.

A few words on the national projection: extrapolation of past trends has yielded fairly good results in aggregate real G. N. P. projections. Obviously, such an extrapolation is based on the premise that the relative trends in causal forces will continue in the future as in the past. Some people have pointed to the significant increases in planned research-and-development outlays—as revealed by National Science Foundation and McGraw-Hill surveys covering the next several years—as a reason for expecting accelerated productivity gains. Actually, the projected magnitudes in relation to probable G. N. P. represent a continuation of the past trends. They are no warrant for expecting accelerated productivity advance; indeed, unless R. and D. outlays (and also educational expenditures) continued to increase more than G. N. P., we might well expect a slowing-down of productivity advance.

Similarly, to project the postwar rate of increase in real product per manhour—which has been greater than the rate of increase over the past 40 years—we would have to project a continuation of the accelerated rate of increase in real capital stocks per manhour. This does not seem unreasonable so long as economic activity generally, and new investment in particular, remain at relatively high levels. But we would also be assuming that innovations are more or less neutral with respect to the relative unit requirements for labor and capital.

Projection of industry rates-of-productivity advance are much more complex than the national projection due to the wider fluctuations from one business-cycle peak to another. Over the cycle, of course, changes in estimated rates of utilization of capacity would be a major element accounting for deviations of output-input changes from their trends.

But looking at intermediate periods—over which plant and equipment as well as labor and materials may be considered a variable factor—there are two main approaches to projecting productivity. The first, and more costly, approach involves technological engineering-type studies. Several studies that have been made indicate that there may be a fairly regular rate-of-diffusion of technique from the most advanced plants throughout the other plants of an

industry.[5] This type of plant study might thus provide a basis for forecasting industry-productivity change. But it would not be enough, since progressive plants would be adopting new equipment and techniques, not previously used, over the forecast period. So "best practices" data would have to be supplemented by engineering data or informed guesses on input requirements of innovations likely to be made. Such studies would obviously be costly.

The second approach is based on statistical correlations. In a Columbia University dissertation, Nestor Terleckyj has found a moderately significant degree of correlation between rates of change in total productivity in manufacturing industries (by using my estimates) and three independent variables: rate of change in output, degree of cyclical variability of output, and research and development intensity. Further correlation analysis of causal factors in productivity change is needed. But in terms of Terleckyj's scheme, the independent variables might be projected as follows: Industry sales and output projections could be tied into real G.N.P. projections through direct correlation, or, preferably (if we had it), through an input-output matrix. The McGraw-Hill survey of planned research-and-development outlays, and expected sales, would be a starting point in assessing changes in the R. and D. ratio.

Productivity projections taking account of factors such as those just mentioned should give better results than mere extrapolation of past trends. The Stanford Research Institute has recently made some broad industry productivity projections for the 1960's by taking into account several causal influences—so if we are still around in 1970 we can test this type of projection. I suggest that, in the meantime, we attempt to improve our knowledge of past productivity movements and interrelationships and, by so doing, increase our power to project and to influence this vital element in economic change.

* * *

[5]See T.Y. Shen, "Job Analysis and Historical Productivities in the American Cotton Textile Industry: A Study in Methodology," Review of Economics and Statistics, May 1958; and Anne Carter, "Innovation and Diffusion, "paper delivered at the Conference on Quantitative Description of Technological Change, April 6-8, 1951 (mimeographed), Social Science Research Council.

SOME FLOW MODELS OF PRODUCTION CONSTRAINTS*

Stedman B. Noble

Logistics Research Project
The George Washington University

After a definition of resource flow models, the study of production constraints is broken into three parts: knowing the effect of a finished equipment upon the economy, recognizing the constraint of the economy, and determining the reactions to be made because of the constraints. Then various methods of analysis are considered: one method being current mobilization planning and another being a model based on the Census of Manufactures. Because of problems encountered in these methods, an abstract analysis is presented that is identified with steps in the desigr of an equipment.

1. INTRODUCTION

This paper considers some of the reasons for the planning of resource flows and offers some techniques that are useful for this purpose. It is sometimes asserted that since our economy is governed by a price system, no planning of the economy is necessary. In fact, planning is going on constantly to prevent unfortunate consequences of constraints on production. Whenever it is desired to have a substantial change in the composition of production, it becomes the task of someone, usually a businessman, to foresee whether any production constraints will occur and, if so, to try to remove them or weaken their influence.

Sometimes, a change in the composition of production requires action by the Government. To show this, it is not necessary to get into controversial political realms, but only to cite the problem of mobilization or semi-mobilization. When the Government wishes to increase the amount of military procurement, or to change from one model of missile to a different version, production constraints of the economy can impose undesired limitations. When changes in the economy are sufficiently large, the industries that produce new equipment may become overloaded with orders, and it becomes difficult to change the capabilities of the plants to produce the new products. If these limitations are foreseen, something usually can be done to increase the productive potential of the critical parts of the economy. This paper considers, in a general way, some of the problems that occur and offers some useful techniques for foretelling constraints that may be imposed by the economy upon military procurement.

This paper is based on the concept of "flows" and considers two different kinds of flow.[1] One of these kinds of flows occurs in the economy itself, and is the flow of commodities and services from one plant to another. What is desired is a model of these flows and in Part 2 some of the problems and possibilities of such models are considered.

*Research was supported by the Office of Naval Research.

[1] Flow models are based on the formalism of directed graphs and the correspondence of these with Boolean matrices. For this mathematics, see Berge, König, and Rosenblatt. For some applications to models of the economy, see Noble, Rasmussen, Rosenblatt [10], [13], and Whitin. For applications to flows of information, see Rosenblatt [13]. Full references are given in the bibliography.

Part 3 considers an organization that is trying to determine the constraints imposed by the economy. For example, a problem might be whether two programs, such as a missile program and a submarine program, conflict with each other by competing for the same productive capacity. In this organization there are flows of information which can be used to construct models of the flows of commodities. This paper considers the purpose and form of models of commodity flows, and the way these models can make use of information that flows in an organization. The models in Part 2 are used to treat some of the problems that occur in Part 3.

2. FLOWS OF COMMODITIES

In this part we consider some available techniques for constructing models of the flows of commodities in the economy. In the first section under this part, a general procedure is described that integrates accounting balances with sequences of flows. The existing data do not allow this procedure, but it is shown that studies of sequences of flows can be used to make selective improvement in accounting models. In the second section, some of the available data are investigated, and a series of models are defined. In the third section we consider the assumptions of a flow model, suggest areas where these assumptions are denied in the models which have been defined, and consider certain specific studies of sequences of flows that are required to adjust the models so as to satisfy the assumptions of flow models.

The Difficulty of Fitting Flows to Accounts

A model of commodity flows may be regarded as a model of accounts. We start with a commodity whose sales are to be predicted, and we must account for these sales by listing the users according to some classification. To obtain quantities, a time period (e.g., a year) must be selected, and the amount of sales going to each user must be estimated. The production (or shipments) must balance with the uses, in that the sum of the production or shipments must equal the sum that goes to all of the users. This is a basic accounting identity that is helpful, if information is incomplete or inadequate, since it defines a definite quantity to be accounted for.[2]

Some of the uses of the commodity may be stipulated in an application of the model: that is, the amount of use may be specified by a policy maker as a starting point for the model. An example would be a commodity that is bought by the military departments, or any item handled by claimant agencies during a mobilization. The other uses that are not stipulated need a further analysis which is done, for each of them, by constructing additional accounts: by listing their sales as going to some classification of users. This procedure, of forming accounts for any sales that are not stipulated, must be continued step-by-step until all sales can be stipulated.

[2]The use of balanced accounts for national planning is an obvious extension of accounting notions that has happened independently numerous times. One existing model of flows, the "input-output" model, is based on a table that balances both along the rows and along the columns. This paper does not develop column balances although they could be introduced when sufficient data were available.

Let us represent this method of model construction by Diagram 1, which,

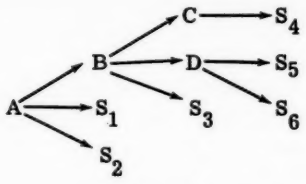


Diagram 1

in the terminology of graph theory, is a "tree." A tree has a source, which in Diagram 1 is industry A, whose output is to be predicted. There are sinks, labelled S, whose amounts are to be stipulated by an administrator when the model is to be used. There are various numbers of arrows leaving any industry; these arrows represent the accounts that are to be obtained. Finally, the distinctive character of a tree is that there is only one directed arrow going to each industry. In other words, Diagram 1 calls for a separate study of flows for each input of an industry.

In practice, data cannot be obtained in the detail required by Diagram 1. Most available data are for entities (industries, firms or plants) which have numerous inputs and numerous outputs. A model can be obtained by connecting the inputs of the entity to its outputs, but such flows are not always "sensible," as is shown by some examples. If the problem is to predict the sales of gasoline engines, and some gasoline engines are sold to be installed in farm equipment,[3] it is necessary to know which types of farm equipment contain gasoline engines in order to know the flows that have occurred. In a model to predict gasoline engines, we should stipulate only the farm equipment that includes gasoline engines.

Now suppose that the problem is to predict the sales of aluminum casting, some of which are sold for use in gasoline engines.[4] However, not all engines use aluminum, so the answer to a more restricted question is needed: which farm equipment contains the type of gasoline engine that uses aluminum castings? This more restricted question requires further information. If the problem is to predict spark plugs, some of which go into new gasoline engines, then the accounts for sales of gasoline engines may be approximately suitable as a step in a model for spark plugs. A more distant relationship, then, sometimes leads to a more restricted set of accounts, but this is not always the case.

One way to be sure that the flows of a model make sense is to construct a model based upon the flows and afterwards to fit the data to this model. Some procedures for accomplishing this are given below. The procedures are difficult and time-consuming, often involving the use of detailed information. In order to conserve effort, it is advisable first to use the available accounting data. After investigating these data, detailed studies may be required to obtain a more accurate statement of the sequences of flows.

To summarize, it is possible to form a model of flows from information on balanced accounts, but when available data are used, the accounts may define flows that do not, in fact, represent flows in the economy. To prevent this, detailed studies of flows are necessary in order to adjust the model so that the flows are meaningful. The next two sections will consider explicit examples of this problem. The next section considers some existing data and

[3] Reference [16], 1954, Vol. I, pp. 210-51 and Vol. II, pp. 35A-37.
[4] Reference [16], 1954, Vol. I, pp. 210-39 and Vol. II, pp. 35A-36.

shows models that can be built using these data. The third section of Part 2 shows examples of incorrect flows based on these data and considers analyses to correct these flows.

Models of Commodity Flows

In this section, explicit models are considered in order to illustrate the data that are available and to examine the difficulties which may occur. The first problem is to determine what the model is trying to predict. One set of models can try to use some balanced accounts that are available in Government statistical reports, and try to predict the same accounts for the future. Two examples that are considered are the allocations of uses of ores given in the Minerals Yearbook, and the purchases of metals that are reported in the Census of Manufactures. These models fit together, giving sequences of flows that start with mining, go through metal refining to metal fabrication and assembly of durable equipment. A model can start at any stage, such as steel mills or metal stampings, and proceed to the higher stages of production that use the products of steel or stamped shapes.

Another way to determine the subject matter of models is to start with industries considered to be of critical importance during a mobilization or wartime period, as determined by the Industry Evaluation Board[5] or by some other agency. Some of these problems would be the mining and metal-fabrication industries considered above and would differ, if at all, by being in considerable detail. The data for such models would rely, in part, on Government statistical reports. In any case, it is important to know what can be done with the Minerals Yearbook and the Census of Manufactures.

The available data give the total production of a material, such as vanadium- or copper-mill shapes and forms, and then lists the users of the material. In the Minerals Yearbook the allocations are quite broad, such as "high-speed steels," "other alloy steels," "chemicals," etc. In the Census of Manufactures, there are allocations of steel, copper, and aluminum, which are given in further detail, such as carbon steel, stainless steel, and other alloy steel. The allocations are the purchases of industries; which gives an allocation to approximately 100 industries. The Census also provides data on total shipments of these metals, but since there are sales that do not terminate in manufacturing (e.g., construction), the allocations do not account for all shipments.

These allocations are only the first stage in model construction. The next step is to follow the flows to the purchasing industries and to form balanced accounts for these industries showing to whom they sell. To obtain these balanced accounts, several sources of data are available. One source is the commodity flow analysis [18] which contains descriptions of which detailed commodities are stipulated and which are not. This permits identification of many industries whose sales are often stipulated. These will be called the assembly industries. For these industries, a graph-theoretic presentation of a model would have the appearance[6] of 2. Diagram 2.

[5]The work of this board is discussed in [17], pages 127-8.
[6]This diagram assumes that repair parts are stipulated. They were treated in this way during World War II, as mentioned in [15, p. 310].

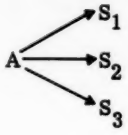


Diagram 2

where A is an assembly industry and S_i is a commodity that is stipulated by a planning agency.

The remaining allocations are those where one industry sells to another. There are allocations in the Census of Manufactures for motors and engines which can yield a model[7] as shown in Diagram 3.

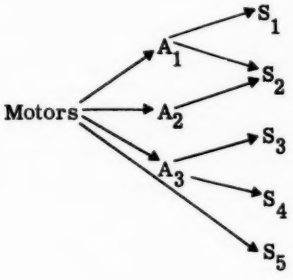


Diagram 3

Now let us consider a model to predict sales of steel. In Table 1 the industries that purchase steel are aggregated into seven categories. The first column shows the number of industries in each category and the next column shows the tons of steel purchased by the industries in each category. The remaining three columns show the amount of steel that is allocated directly, or indirectly, or not at all. The categories of industries are defined to show some types of diagrams that are required. The diagram for steel going to the assembly industries is given by Diagram 4, which is similar to Diagram 3.

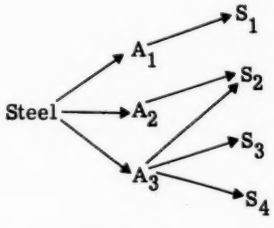


Diagram 4

[7]Some care must be taken to obtain such a simple model. Motors and engines sell to the part of the pump and compressor industry, for example, that makes producers' durables rather than components. Furthermore, repair parts must be stipulated.

Most containers are tin cans used for food, as seen in diagram 5.

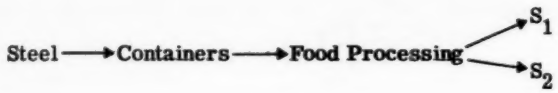


Diagram 5

For the next three categories of industries, allocations were made by noting that commodity designations, such as motor-vehicle stampings or locomotive and railroad-car springs, implicitly define approximate allocations. These allocations almost always go to some of the assembly industries, as shown by Diagram 6. These are the allocations that are called "indirect" in Table 1.

TABLE 1

Allocations of Steel Inputs by Type of Industry[a]

Type of Industry	Number of Industries	Inputs[b]	Allocated to End Products		Not Allocated
			Directly[c]	Indirectly[d]	
Assembly	71	21.2	21.2		
Containers	4	4.9	4.9		
Engines and Motors	3	0.7		0.5	0.2
Components	20	5.2	0.3	2.3	2.6
Metal-Working Operations	18	12.9	7.6	2.7	2.6
Metal-Working Overhead	6	0.4			0.4
Mixed Type	7	0.6			0.6
TOTAL	129	45.9	34.0	5.5	6.4

[a] Developed from the 1954 Census of Manufacturers as discussed in [8]. Units given are in millions of short tons.
[b] Inputs of steel are available for only 99 of the 129 industries.
[c] Derived mainly from the commodity flow analysis.
[d] These allocations are based on the commodity names or are given in the Census of Manufacturers.

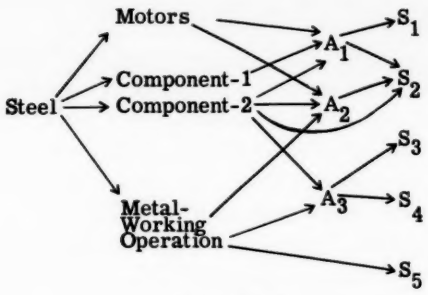


Diagram 6

Table 1 shows that Diagrams 4 and 5 account for over half of the steel, and that, together with Diagram 6, 85 percent of the steel total is accounted for.[8] Much of the remaining 15 percent has been allocated in the interindustry table developed by the Bureau of Labor Statistics [3] based on the 1947 Census of Manufactures. In this table, the sequences become longer, and, if all allocations are considered, the diagram would become complicated and involve cycles (i. e., feedback).[9] However, if the "small" coefficients are ignored, then the cycles disappear and a flow very much like, but slightly longer than, the previous diagrams is obtained in Diagram 7.

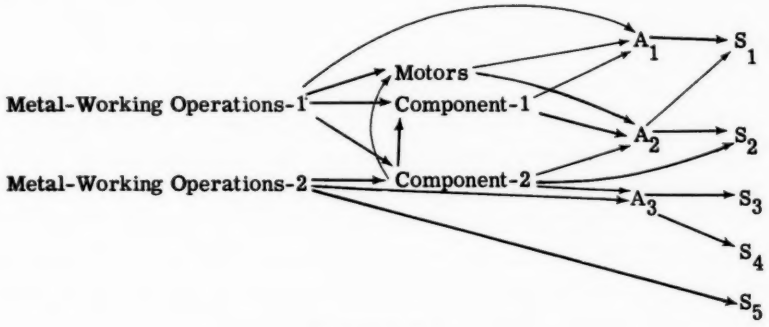


Diagram 7

Steel sells to every category in this diagram except the stipulated activities.

Models similar to these can be constructed for other metals or for kinds and shapes of steel. In turn, it is possible to develop models for blast furnaces and iron ore, or for some of the sales of manganese, chromium and other alloys, by extending the above models. Starting with an allocation of chromium to stainless steel, the model for stainless steel can be used to continue these flows. For other metals, such as tungsten, the problem is more difficult since the Census data allocates simply all alloy steel, and does not designate it by type of alloy. We will return to this problem, but before doing so, let us consider the assumptions underlying the construction of such numerical flow models.

---

[8] In [8] there is a more complete discussion of the methods used to produce Table 1.

[9] Flow models of the kind given here (where the flows are in one direction) can be translated into triangular matrices and are easy to compute. When the graph involves small amounts of "feedback", there are efficient methods of computation upon the corresponding matrix.

The Assumptions of a Flow Model

Several assumptions are made about the flows that are diagramed in the previous section. First, it is assumed that these flows can be represented by numerical input coefficients. In the flow of pig iron from blast furnaces to steel mills, the input coefficient is 3/10, which means that out of every dollar of sales of steel, thirty cents will be spent for pig iron. The first problem in the construction of a flow model is to test the available data to determine whether the input coefficients are stable over time.

The second assumption stems from the fact that certain flows are connected; that is, one arrow goes into an industry and another arrow flows out of the same industry. It is assumed that there is a relationship of "cause" and "effect" between the sectors that are thus connected. If pig iron and ferromanganese flow into steel, and steel flows into automobiles and containers, it is assumed that some pig iron and ferromanganese production occurs because of the demand for automobiles and containers.

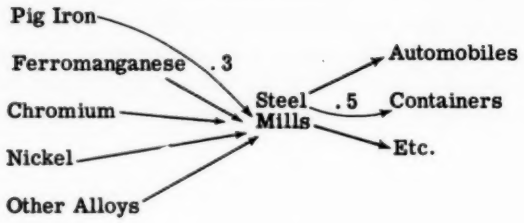


Diagram 8

Actually, a stronger version of this assumption is required. When flows are connected, then the input coefficients associated with the flows are multiplied times each other. The basic assumption of such a flow model is that this multiplication yields a predictive functional relationship If the input coefficient of pig iron into steel is 3/10 and the input coefficient of steel into containers is 5/10, then the production of containers requires a production of pig iron of 3/10 times 5/10 or 15/100. That is, if plants that make containers also produce their own steel,[10] then they would spend 15/100 of their costs for pig iron.

To restate this last assumption: a model gives predictive results only if the inputs are independent of the composition of the output.[11] Independence exists for the flows of pig iron and manganese into steel mills but not for chromium and nickel because they go into stainless steel but not into carbon steel. A purchaser of carbon steel does not generate requirements for production of chromium, nickel, or for most other alloying metals.

How do we tell whether a flow model satisfies the assumption of independence? It is necessary to consider the sequences of flows that are generated in the model structure, and to think through whether they define predictive functional relationships. Two subject-areas, metals and component parts, where this needs to be done will be discussed.

On the subject of metals, an inspection of the available data shows industries which probably do not satisfy the assumption of independence. Table 2 lists the ten industries which

[10] This statement is an economic interpretation of mathematical elimination, which is a method that may be used to solve a model.

[11] The assumption of independence has been discussed in a somewhat similar way by Wold and others. A more comparable assumption of independence is made in the probability theory as is discussed by Kac, page 10. These references are from D. Rosenblatt.

TABLE 2

Metal Purchases of the Industries Purchasing Large Amounts of Aluminum[a]

SIC No.	Purchasing Industry	Percentage of the Sales of Aluminum	Metals Purchased			Aluminum as a Percentage of the Three Metals
			Aluminum[b]	Steel[c]	Copper[d]	
3442	Metal doors, sash & trim	11.2	224,734	796,226	1,640	22
3721	Aircraft	9.3	186,403	127,542	8,502	58
3717	Motor vehicles & parts	9.1	183,105[e]	18,890,148	210,357	1
3463	Metal stampings	8.4	167,627	4,546,406	64,911	4
3497	Metal foil	8.1	161,303	-[f]	-[f]	100[f]
3444	Sheet-metal work	6.4	127,789	2,277,036	8,115	5
3729	Aircraft equipments n e c	4.9	97,477	401,250	15,146	19
3585	Refrigeration equipment	4.6	91,864	1,609,062	84,119	5
3715	Truck trailers	2.5	50,598	318,056	2,011	14
3661	Radios & related products	2.2	43,876	285,534	84,737	11
	71 other industries	66.7	1,324,311			
	Total	33.3	677,222			
		100.0	2,001,533			

[a]Source: 1954 Census of Manufactures. The Census data are converted from short tons to 1000 lbs, the unit used here.
[b]Aluminum and aluminum-base alloy castings and mill shapes and forms. Also included are refining shapes and scrap purchased by the industries. SIC 34 through 37.
[c]Steel mill shapes and forms. Iron and steel castings are not included.
[d]Copper and copper-base alloy castings and mill shapes and forms.
[e]Of this, 145,925 is aluminum castings.
[f]Small amounts of steel and copper may be bought, but the figures are not available.

purchase the largest amounts of aluminum. All but one of these industries purchase large amount of other metals, usually steel. This raises the question as to whether the metals are functionally dependent upon every sale of these industires, a problem which is represented in Diagram 9. The numerical aspect of the independence assumption is particularly

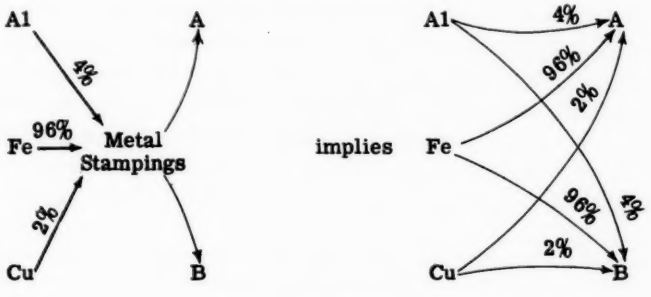


Diagram 9

difficult to accept.[12] For the steel industry, the problem shown by Table 2 is not serious because its major sales go to industries that primarily buy steel. In the case of aluminum, copper, and the additive and alloying metals, however, special studies are required to determine whether or not the assumption of independence is satisfied.

Detailed studies can be made by using technological information, and one example is given for stainless steel[13] showing four of the over thirty kinds of stainless steel, most of which have different metal composition, different properties, and different uses. In the case of alloys, approximate quantities can be obtained on the composition of the inputs, but it is not

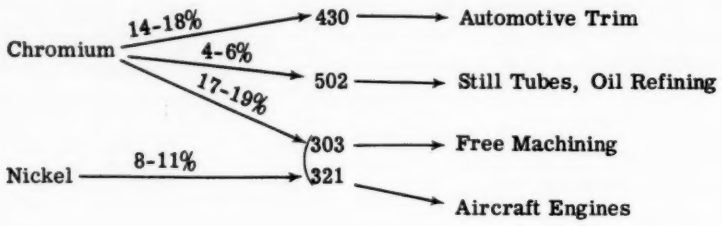


Diagram 10

possible to obtain a complete list of outputs or to obtain coefficients for the output flows. What is more typical is to discover some use or function that is served by the alloy. From the functions and the properties of the alloys, it must be estimated what objects would include alloys with a defined input content. After completing such flow diagrams, we can return to the more aggregate data of the Census of Manufactures which provide two kinds of control totals: the allocations of stainless steel to each industry, and the inputs of chromium and nickel to the steel industry, much of which goes into stainless steel. Using these data as control totals and the flow diagram as a pattern of causation, the amount of chromium and nickel going to each user of stainless steel can be estimated.[14]

There are some cases in which flow models may be required for accurate models of component parts. An efficient way to analyze these models is to consider the concept of inputs that are embodied in the finished product. For example, a tractor which has an engine and wheels, a clutch, a seat, etc. Engines have oil pumps, clutches have gears that are machined, and all of these are made of metal. This type of information can be obtained from repair manuals that exist for most of the complicated equipments and assemblies that are mass produced. The embodied inputs in a finished equipment imply a flow of goods that occurred in the course of production of the equipment: steel went into oil pumps that went into engines that went into tractors.

Once the embodied inputs of some equipments are known, equipments that do not have the same embodied inputs can be determined. Consider whether the inputs go into every equipment or not. Every tractor has an engine, but only some tractors have windshields. Other embodied inputs compete with each other. An example is that of diesel engines and steam

[12] Most of the industries in Table 2 produce commodities that use different amounts of each metal, and these commodities go to specific users. Examples: aluminum doors in house trailers, or the large ratio of steel to other metals in a helicopter. A model using these industries as they stand would not define the predictive functional variables for aluminum.

[13] See [1]. page 34.

[14] The Defense Materials System, which provides data for military programs, distinguishes nickel-bearing and other stainless steel.

turbines in ships. Another example is that of gasoline engines and electric motors that go into power lawn mowers.

A study of the embodied inputs is also necessary in order to estimate which metals are purchased by the assembly industries. The problem is that the industries called assembly industries do more than assemble: they produce a large number of components in addition to purchasing some components. When these industries purchase alloy steels, it is necessary to know which components they produce in order to know where the steel is being used. This requires a flow model that lists all of the major components of an assembled equipment, which can be used to estimate the components that are purchased and those that are self-produced. Then metal and alloy composition of the self-produced components can be estimated.

Let us review this method of constructing a logically coherent flow model. First, the available data should be tested for "constant" relationships. Then areas where the assumption of independence may be violated, such as metals or certain components, should be identified. In the case of metals, the existing data reveals places where there may be a problem of accuracy. Using technical sources, flow models are built to reveal the detailed flow patterns. These flow models are used to estimate the quantities implicit in more detailed models.[15]

It should be noted that a model based on flows and graph theory can be translated into a corresponding matrix model and computed using methods of numerical analysis that apply to matrices. A graph corresponds to a Boolean matrix and where there are linear functions on the graph, these become coefficients of a matrix. This correspondence between a matrix and a graph is discussed in Rosenblatt and in Noble [6].

Summary

There is no theoretical problem concerning how to turn balanced accounts into a model that describes flows in the economy. The problem occurs, however, because the available data are collected in a uniform grid that does not provide the basis for accurate models for each specific problem that may be encountered. The problem is lessened with more detailed data, but many of the questions asked lie beyond the scope of the existing data

A study of some of the existing data has been made, and it is possible to assemble fairly crude accounting models. The major part of these models can be assembled fairly routinely using several different data sources. There are areas, however, especially for some metallic components, where it is difficult to make allocations.

The available data can provide models that are suitable for the allocation of steel and its derivative inputs, such as manganese, pig iron, and iron ore. The models, however, are not sufficiently accurate for models of the other metals, particularly for alloying materials. Such metals require the identification of major areas contributing error, the construction of flow models to elucidate the detailed causation, and an estimate of the quantity of flows that will give a more accurate prediction. The detailed flow models are constructed by using the concept of embodied inputs and various technological information.

[15] The authors of literature on the "input-output" model are aware that models may generate spurious relationships: the problem is called "the aggregation problem" and the suggested solution is the development of a more detailed matrix model. After using the most detailed accessible data, however, substantial spurious relationships still remain. The present paper suggests that the solution is not matrix disaggregation, but the construction of structurally coherent flow models, which can be made using auxiliary data that are not used in the preparation of "input-output" models. For further discussion, see Noble [7, 8].

3. FLOW OF INFORMATION

This portion of the paper considers an organization that is trying to determine the production constraints imposed on it by its suppliers. In an organization searching for production constraints, there is usually a flow of information from one part of the organization to another which can provide data for models of the economy. Even if a model is not constructed, the concepts of production models can provide a framework suggesting which information should flow to a person who is to make the decisions.

This part of the paper relates to the previous part in the following way. By considering constraints, it goes beyond the problem of Part 2. In considering models, it extends the previous part by dealing with more complicated models which incorporate the models of Part 2. When the data for such models are considered, it can be noted that the previous part has introduced the kinds of data used for either kind of model, and it is important to know where these data are located in an organization.

This part is not a definitive discussion of a theory of organization. Neither is it a study of the methods that are being used or should be used by an existing organization. Rather, it extends Part 2 to show in general the role of models in an organization. This is done by distinguishing three subproblems, each of which is typically the concern of a different group of people in a large organization. The subproblems are the strategic or decision problem, the design problem, and the production problem. Each of these will be dealt with in a separate section of the paper.

The strategic or decision problem (the subject of the first part of Section 3) is to choose among the conflicting claims of (1) the properties of an equipment, (2) the time it takes to get into production, (3) the rates of production, and (4) the cost. Each of these may vary when there is a change in the others. To define the choices that are involved, it is necessary to have information on the way that one variable, such as cost or time of production, is affected when another variable, such as a property of the finished equipment, is changed.

The design problem is dealt with in the second part. A design translates the requirements of an equipment into documents that can be used for production and purchase. The design process can provide information on the sequences of flows that are required for the production of an equipment.

The production problem is the topic of the third part. It is at the point of production that the actual capacity of the economy is determined. The problem is to determine what information about production needs to be assembled to obtain a picture of the constraints provided by the entire economy.

The Problem of Strategic Decisions

Before the nature of the strategic decisions that affect the economy are considered, let us consider why a feasibility test is an incorrect way to handle strategic decisions. The feasibility test starts with a program which is used to develop a set of strategic decisions as to the number and kinds of equipments and assemblies desired during a mobilization period. A study is made which determines, after considerable time, whether the program is feasible. In the meantime, new information may be obtained by the strategist that causes him to change his program. Thus, the feasibility test provides training for an organization in the conduct of detailed studies, but it often does not inform the strategist as to the feasibility of his programs. The number of each equipment to be produced is one of the last things that should be decided

because it is the kind of thing that is most apt to change. The relevance of different equipments changes substantially depending on the kind of war, the opponent, the location of the war, the state of technology, etc. Judgement about some of these factors changes frequently, and the changes often permanently eliminate the need for certain equipments and make other equipments critical.

Consider an emergency entailing the working of a limited number of production decisions. The emergency stems from new information about the strategic situation, so the decision-maker has new priorities about equipment and wishes to change some of the existing decisions about production. To do this, three questions become important: (1) For a given equipment, such as a gas mask, at what rate can the economy produce it? (2) What changes must be made in the economy to get the required production? (3) What different equipments compete with each other for the same production capacity? If there are answers to these questions, the decision-maker can know whether he can expand the production of equipments that are strategically desirable, what actions he must take, and the cases where production of one equipment can be obtained only by cutting back the production of other equipments.

A minimal need is a map of the flows in the economy: a flow model of the economy. Such a model starts with a list of the finished equipments and assemblies to be produced by the economy. As a guide, rough amounts for each equipment can also be stated, but these are preliminary since the actual amounts are to be decided later in an emergency, partly with the help of the model. The strategist is responsible for deciding what equipments are to be studied, and the rough amounts of each. Just as more designs will be prepared than are to be used, so more equipments will usually be analyzed than will be produced during any actual mobilization period.

What degree of accuracy is required in making this model? To answer this, we must distinguish between the strategist and the entire organization. The organization must make accurate decisions based upon very detailed information and analysis. The decisions made by the strategist are a limited part of the entire set of these organizational decisions. From the point of view of the economy, the main decisions of the strategist are the number of specific equipments to be produced, and to make these decisions correctly, he requires only a limited amount of information about the economy. Moreover, before a specific decision is to be made, more explicit information can be collected and advice can be obtained. This means that the flows need to be known much more accurately than the quantities. The flows should tell precisely what relationships are involved, so that specific questions can be defined when further analysis becomes desirable.

Design Information

A design involves translating certain requirements of an equipment into a set of documents that can be used for procurement and production. A design is not finished unless the designer determines, explicitly or implicitly, that the various components could be produced and the various processes implicit in the design could be executed. In other words, a design is supposed to be feasible.

What the strategist wishes to know is not only whether the design is feasible, but also the constraints of the economy which determine the length of time for production to start and the effective rate of production. To discover this, the knowledge of the designer must be objectified so that the decision-maker can use it and relate it to information obtained from producers.

What needs to be objectified is the pattern of sequences of flows of commodities that go into the finished equipment. [16]

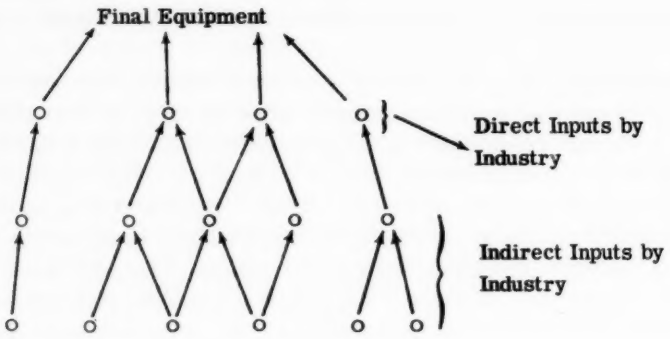


Diagram 11

To see more precisely what is involved, consider Diagram 11. Consider planning that starts with the designation of a final equipment that probably should be produced. The next problem is to know the direct inputs of the equipment,—a problem which can be solved perhaps better by the purchasing office than by the designer, whose data are generally too detailed. There are two reasons why this detail is not necessary. First, the purpose of the model is to locate critical industries where an industry is a collection of plants. Thus we can aggregate all items bought from the same plant or the same kind of plant. Second, many industries may not be sufficiently critical to require a detailed study, so we can focus on the remaining industries that are more important to study.

In obtaining data on indirect inputs, much more precise information is required, so it is necessary to use design information. Examples of what needs to be known are the parts within a purchased engine, or the metal and coating used in purchased wire. Here the specifications for the parts need to be known. Actually, any one design process may make use of components and materials without analyzing them further, and thus without knowing the flows that precede them. For example, a designer may use a vacuum tube without knowing either the kind of metal that is used in it or the processing that must be done to this metal. It may be necessary to go to other flow sources, such as handbooks, or to other designers in order to continue the flow analysis.

An analysis by a designer is similar to the study of embodied inputs that was discussed in Part 2. However, it is possible to extend the concept of embodied inputs to include any inputs that can be identified from knowledge of a specific equipment. This identification can occur through inspection, disassembly, chemical analysis, studying manuals, or studying design documents. We can extend the concept to include flows of any inputs that can be inferred.

[16] This is a production planning method as applied to large portions of the economy. This method has been formalized by Vazsonyi in his discussion of "Gozinto" graphs. Another name for this method is a "bill of materials" method and such studies are now made by the military departments.

Some components can be known to require machining on various kinds of machines. Some metal refining can be known to require furnaces which in turn use carbon or graphite electrodes that have to be replaced. The information that is required concerns indirect inferences of various kinds.

Besides knowing the indirect effects, we are also concerned with knowing the substitutions that are possible between inputs. Here again, the existing engineering knowledge is voluminous, so any obvious substitutions should be developed by more accessible information sources and only special problem areas need to be put to a designer.

A major reason for obtaining information from the designer is that he is responsible for the detailed design decisions. The designer must pass on the possibility of substitutions and be prepared to carry out any redesign that becomes necessary. Also, there are constant changes in technology and information about materials which require the up-dating of a model. In this sense, the designer works with commodity flows before they occur, so he is a logical source for information to update a flow model.

Capacity

The capacity of the economy is ultimately determined by the response of the production managers and foremen of plants, and several kinds of response can be distinguished. Each plant or department decides, when confronted with a production problem, whether it is able to do the job. If it accepts the order, then the plant estimates the time required to produce the order. Finally, if an agreement is made,[17] there is an actual time when the production occurs. These decisions are affected by changes in circumstances outside of the plant. First, the acceptance of an order and the timing of it are affected by the existence of other orders and by the likelihood of getting other orders. Second, these responses can be changed somewhat by obtaining new capital equipment and personnel. The responses of a plant are affected by various limitations or bottlenecks, so the response may be altered by relaxing these limitations. The limitations could be shortages of workmen of various skills, equipment of various kinds, rates of processing purchase requisitions, or even the physical limitations of speed or size. Sometimes the limitations can be relaxed so that a plant increases the kinds of orders it can take. In any case, the rate of production can be increased by relaxing certain specific limitations.

What the planner wishes to do is to influence the plants in the economy so as to get more production of the kinds desired by the planner. He can do this by affecting the orders accepted by the plants and by changing the material and capital equipment going to the plant. The first method of change is to influence the expectations of the plants. The plants have an opinion, from their knowledge of the world situation, as to the demand for their products. When the world situation is confused, a planner's main task is often to define the situation so that the plants can prepare for new circumstances. A mobilization is such a period: when a statement by a planner will allow the plants to know that war production is going to increase. Of course, the planners must also follow through on their statements: that is, the orders to the plants must change to wartime orders. A planner can even cancel existing orders if he is

[17] The agreement depends, of course, upon the price. During a mobilization period, however, the price becomes less important to both parties. The Government becomes less cost-conscious and the company may be confronted with stop-orders, controls and the possibility of renegotiation of contracts; all of which help to cause the acceptance of orders to depend more upon physical capacity and less upon costs of production.

sufficiently sure which plants are going to be of critical importance. The other lever of change available to the planners is to make material, capital equipment, designers, workers, etc. available to the plants whose capacities are of critical importance.

The planner can evaluate the importance of end items, not of plants; so to take action regarding plants, he must know the relationship between plants and end items. This requires four distinct steps. The first step is to start with the end items and find the sequences of commodities that go into them. This has already been discussed in the previous sections. The second step is to determine all the possible plants that can produce any commodity that is suspected of being in short supply. There is some variation in the list of possible plants due to the changes in incentive and changes in capital equipment, but the variation is fairly narrow. The third step is to determine the possible production rates, that is the capacity, for these plants; and furthermore, to decide what other commodities compete for the same portion of the plant capacity. When there are no other commodities competing for the plant capacity, and no uses of the commodity except the uses already analyzed, then we have a simple model: the capacity of this group of plants becomes a capacity of the economy upon the end equipments. This can be the case for many commodities bought by the military during a mobilization.

If there are other uses for the commodity under analysis, or other commodities that can be made in the same plants, then the fourth step is necessary. A model must be constructed that allocates the commodities to the end uses that require these commodities. The construction of such a model is discussed in Part 2 above.

The entire sequence is given in Diagram 12. The first step is to study specified end products in detail, the next two steps are to determine what commodities compete for the same production space by determining what plants can make a given commodity, and which further commodities compete for the same productive capacity of each plant. A fourth step is to trace the additional uses of the same or other commodities to find what end uses call for the production of these commodities.

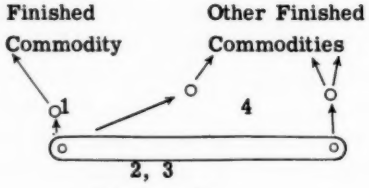


Diagram 12

This sequence defines what finished equipments are competitive with each other by way of the capacity of a group of plants. There remains the problem of determining what are the most likely points of capacity for any possible mobilization program. This can be probed before an emergency occurs by attaching numbers to the flows on the graph, and deriving approximate quantities for capacity.[18] Such analysis should locate a limited number of potential bottlenecks which can be "scheduled" during a mobilization either by requiring frequent reports on the status of production or by changing the requirements for competing programs.

The commodites that are stipulated in one model can become dependent variables in another model. This can be true of consumer goods such as gasoline for the trip to work or

[18]After formulating a quantitative model, these problems remain: the testing of available data to find stable functions, and the development of efficient computational methods.

new housing in a "critical area." Such an extension of the model is also possible for capital equipment which can change and expand the capacity of "critical" parts of the economy.

There is a limited capacity to produce new capital equipment, but a model could suggest where this should go to help loosen critical bottlenecks. Any detailed allocation of new capital equipment, or redistribution of existing capital equipment or highly skilled personnel, can be defined in a model but a very detailed and accurate model would be necessary. This is an area, common in administrative problems, where the overall view can define the problem, but actual decisions are best delegated to those who can locate and take into account detailed information.

4. SUMMARY

The model of the last part is idealized and does not discuss important operational problems or efficiencies. The purpose of the discussion is to extend the discussion of models in Part 2, showing the kind of model that is required by an organization doing mobilization planning, and suggesting how the organization might develop such a model. The discussion purposefully refrains from discussing the military-civilian aspects of planning.[19]

The model of the last part starts with a list of the finished equipments and then analyzes the commodities needed to produce each of these. These inputs can be obtained from various sources, but essentially design handbooks or designers can provide necessary detailed information. The commodities must be related to the plants that can make them, and we need to know the capacity of these plants and the other commodities that compete for the same capacity of the plants. These other commodities must be studied to determine in what end items they are needed. Once this is done, a decision is possible as to which end items are the most important.

The last step, of relating commodities to their end items, is the same problem as the models in Part 2. These models are based upon accounting data of the kind available from Government sources. It is not difficult to obtain models for most portions of the economy, but it does require considerable care to develop accurate models. The problems of accuracy often require a detailed analysis of flows and then an estimation of coefficients that apply to these flows.

The state of the art of model construction is such that fairly precise information can be obtained on the existence of flows of the economy, but only imprecise quantities can be specified. This fits rather closely to the needs for models during a mobilization. The decisions of a mobilization require a knowledge as to what equipments can be produced, and what end equipments and components compete with each other. Typically, once this much is known, the correct decision is fairly obvious or else a quick and detailed study can provide the answer. A decision maker may want approximate quantities in order to weed out minor effects, but this amount of numerical accuracy probably is possible at the present time.

I conclude that models of the economy can be obtained which will make more information available to the high levels of decision making and will define more sharply the choices that are available.

---

[19] The problem of making decisions and obtaining data to help these decisions can be analyzed as if there were one organization, and then the decision-making and data collecting can be delegated to suitable agencies.

ACKNOWLEDGMENTS

This paper stems from the suggestions, discussion, support, and criticism by a large number of people associated with the Logistics Research Project, including J. E. Hamilton, W. H. Marlow, R. E. McShane, David Rosenblatt, Frank Segel, Henry Solomon, and A. Temple. None of these are responsible, however, for limitations or omissions, in this paper.

BIBLIOGRAPHY

[1] American Society for Metals, Metals Handbook, 1954 Supplement, Taylor Lyman, editor, Cleveland, Ohio 1954.

[2] Berge, Claude, Theorie des Graphes et ses Applications, Paris, Dunod, 1958.

[3] Evans, W. Duane and Marvin Hoffenberg, "The Interindustry Relations Study for 1946," Review of Economics and Statistics, Vol. 34, 1952, pp. 97-142.

[4] Kac, Mark, Statistical Independence in Probability, Analysis and Number Theory, Mathematical Association of America, Wiley & Sons, N. Y., 1959.

[5] König, Dénes, Theorie der Endlichen und Unendlichen Graphen, Akademische Verlagsgesselchaft, MBH, 1936. Also, Chelsea Publishing Co., N. Y., 1950.

[6] Noble, Stedman B., Measures of the Structure of Some Static Linear Economic Models, Logistics Research Project, The George Washington University, 11 August, 1958.

[7] _____, Structure and Classification in Resource Flow Models, Logistics Research Project, The George Washington University, 19 May, 1959. Submitted for publication.

[8] _____, Resource Flow Models with Application to Metal Fabrication, presented before the Econometric Society, Dec. 1959; abstract forthcoming in Econometrica.

[9] Rasmussen, P. Norregaard, Studies in Intersectoral Relations, North-Holland Publishing Co., Amsterdam, 1956.

[10] Rosenblatt, David, "On Linear Models and the Graphs of Minkowski-Leontief Matrices," Econometrica, 1957, pp. 325-338.

[11] _____, "On Aggregation and Consolidation in Finite Substochastic Systems," Annals of Mathematical Statistics, abstracts, Vol. 28, 1957, pp. 1060-1061.

[12] _____, "On the Graphs and Asymptotic Forms of Finite Boolean Relation Matrices and Stochastic Matrices," Naval Research Logistics Quarterly, Vol. 4, 1957, pp. 151-167.

[13] _____, "On Some Aspects of Models of Complex Behavioral Systems," in Information and Decision Processes, edited by Robert E. Machol, McGraw-Hill, N. Y., 1960.

[14] U. S. Bureau of Mines, Minerals Yearbook, Vol. I, Washington, D. C., Government Printing Office.

[15] U. S. Bureau of the Budget, The United States at War, Washington, D. C., Government Printing Office, 1946.

[16] U. S. Bureau of the Census, United States Census of Manufactures, (exists for 1947, 1954, 1958), Washington, D. C., Government Printing Office.

[17] U. S. House of Representatives, Ninth Annual Report of the Activities of the Joint Committee on Defense Production, 86th Congress, Second Session, House Report #1193, Washington, D. C., Government Printing Office, 1960.

[18] U. S. Office of Business Economics, National Income, 1954 edition, Washington, D. C., Government Printing Office, 1954.

[19] Vazsonyi, Andrew, "The Use of Mathematics in Production and Inventory Control," Management Science, Vol. I, part 1.

[20] Whitin, T. M., "A Note on the Existence of a Static Equilibrium in the Leontief System," Weltsirtschaftliches Archiv, Band 69, 1952, pages 188-196. Reprinted in Economic Activity Analysis, editor O. Morgenstern.

[21] Wold, H., "Ordinal Preferences or Cardinal Utility?" Econometrica, Vol. 20, 1952, pp. 661 ff. See also notes or articles by G. L. S. Shaskle, L. J. Savage, Alan S. Manne, A. Charnes, Paul A. Samuelson, and E. Malinvaud, same reference, pp. 661-679.

* * *

PANEL 3a
MACROSIMULATION IN LOGISTICS RESEARCH

THE USE OF MAN-MACHINE SIMULATION FOR SUPPORT PLANNING

Murray A. Geisler

The RAND Corporation

Support planning under present conditions of rapid exploitation of technological advances to maintain the most advanced weapon inventory creates many new stresses for planning the support of these new weapons. In addition, support planning has always been incomplete in that there has not been a systematic way of adjusting plans to reflect operational experience.

The second major experiment of the RAND Logistics Systems Laboratory has produced techniques of support planning for new weapons which may represent major advances in such activities. This paper describes the conditions under which support planning must be performed in the present and future military environment and the experience of LP-II in performing such planning for an ICBM weapon of the 1963-65 time period, as an example.

THE ENVIRONMENT

The environment that confronts the military planner is marked primarily by an accelerated development of new weapons. Premium is placed in military planning upon the possession of the most modern and advanced weapons that the country can develop. This desire for rapid weapons progress also requires that the time needed to make these weapons operational be as short as possible once they are developed. By an operational unit, we mean one that has the capability and support needed to perform its operational functions. Such an operational unit would have all the specified resources and the necessary training for operating the weapon effectively as well as maintaining the weapon in an operational condition over the long run. Thus, a basic prerequisite for a weapon system to move from the research and development stage to the operational status is the provision of the specified support in terms of people, equipment, materiel, and training. The ICBM program is a good illustration of this new situation in which there has been an intensive effort to develop the weapon and to produce the support concurrently, so that there will be a minimum delay in its being operational once the missile attains the desired level of reliability.

Because of these rapid technological advances, drastic changes in weapon characteristics and in the environment created by the use of these weapons can be expected. Thus, the task of the military planner, and in particular that of the support planner, is made very complex and difficult. He is required, on the one hand, to accelerate the pace of his support planning in order to make the new weapons operational in as short a time as possible; and on

the other hand, he is compelled to deal with radically different weapons so that the experience available to him from earlier weapons is of little application or even possibly misleading.

The environmental requirements have not been as demanding in the past. Previously, there was usually a testing or proving interval planned between weapon development and operational status in which the proving of the weapon provided experience and data from which the bulk of the support requirements were determined. Thus, it was possible for the support planner to wait for actual experience upon which to project a large segment of the support requirements, whereas in the future this opportunity will not be available. In addition, in past programs, such as the B-47, F-100, etc., large quantities of aircraft were produced, so that it was possible to accumulate operational experience on the earlier aircraft that could be used in determining the support for the later production. In the newer weapon systems, this will not be possible, since the trend is to small numbers produced, so that the production virtually terminates at or about the time the first weapon becomes operational. Thus, the prospect is that the difficulties of support planning will be intensified in that the useful life of new weapons will be increasingly curtailed, and the differences between successive weapons will become even more radical.

In addition, support planning must take place under the usual conditions of uncertainty. The pace of weapon development is, at best, an educated guess. One subsystem that was expected to do well turns out to be a failure. A new breakthrough in propulsion design is achieved. The testing program runs into unexpected difficulties. High reliability of the weapon is often a hoped for, rather than an achieved goal. Operational concepts change and conflict with budgeting constraints, and the enemy always complicates the situation with unexpected strategies. All of these types of elements fundamentally affect the support planner and his techniques for support planning.

Planners must be able to live with this future uncertainty. Techniques which provide flexibility and responsiveness in support planning are the means by which such uncertainty can be managed. In support planning, this would include the ability of the planner to revise his plan rapidly and to examine alternatives. It also encourages him to reduce the lead time required for carrying out the implications of the support plan in terms of production, training, etc., so that he can delay until the last possible moment taking the actions implied in order to use the latest planning data. On the other hand, the reduction in time required to make new weapons operational helps in dealing with this future uncertainty because it requires a shorter projection into the future; and it permits the military to capitalize more rapidly on the weapon, and therefore to maintain a closer correspondence between the pace of development and the pace of weapon use.

NATURE OF SUPPORT PLANNING

The function of a support plan is to define the resources required to carry out a given operational plan. It must be internally consistent so that the individual activities that rely on this plan deal with requirements that are based on common assumptions and integrated requirements.

Under the new need to provide support as soon as a weapon is developed, support planning must begin very early in the weapon system life cycle; in fact, support planning must begin as soon as the broadest information is available as to the concept of weapon operation and design. The support planning thus necessarily begins on the broadest level, and it must

successively proceed to greater amounts of detail. Thus, the initial support planning first considers the general characteristics of the support structure required by the proposed deployment of the operational force, the number of weapons to be placed at each operating base, and the weapon reliability, operating concepts, and repair characteristics.

These broad considerations of support must be rapidly followed by a more detailed statement because the production-training-procurement agencies must be provided adequate lead time to produce the desired support at the required time. The lead times for obtaining these resources are such that highly definitive resource estimates and procurement actions must be accomplished before even the first weapon is operational. In other words, the stress has not only been placed upon having support at the earliest time that a weapon can be made operational, but also in producing detailed estimates of such support requirements even earlier in the weapon development cycle.

Further, the estimates of future support requirements will have to cover a wider range of resources in the new weapons. They will include more emphasis on ground support and facilities equipment, in addition to personnel and spares, and they will have to be integrated estimates accomplished through joint planning. This is because there is much greater specialization in equipment and functions in the new weapon systems. Standardization is very difficult because of the time stress and the limited production quantities. Since such resource requirements hinge directly on weapon characteristics and operational demands, the absence of experience will require the planner to draw upon any relevant planning and technical information he can use. The only data that are available to the support planner under these circumstances are engineering judgments, development data, particularly for components and subsystems, and past experience on similar equipment and functions, ranging from highly relevant experience to highly tenuous assumptions. The ability of the support planner to digest these data in a valid and systematic way will be a fundamental determinant of his efficient use of support resources. This also includes the need to adjust the support plan to reflect the influence of changes in technical data, program data, etc., as experience unfolds and programs may change.

IMPLICATIONS FOR SUPPORT PLANNERS

Thus, whereas previously the support planner could get by with broad and consequently uncommitting estimates of support until he could collect experience, under the new environment, the support planner is required very early during the development period to produce detailed estimates of support requirements. For personnel, this means stipulating manpower requirements by specialty and skill. For equipment, this means stating the equipment requirements by specific class or category. For transportation, the requirements must be given by type of vehicle, and for spare components it must be by type of component.

Clearly, the technique used to produce such estimates must be detailed enough so that the impact of operational reliability and maintainability assumptions upon the requirements for a particular component or a particular skill can be measured. Also, the impact of specific support limitations on operational capability must be measurable in developing the support plan and in selecting feasible operational assumptions. If the support planners cannot provide the relevant requirements in adequate detail because of limitations in their techniques, then the procuring or training agencies will make assumptions about what these requirements must be in order to provide themselves with the relevant lead time for producing the resources. This practice leads to an unbalanced and uneconomical production of resources.

In addition to imposing a requirement for more detailed planning early in the weapon system cycle, the increasing uncertainty about the future, the more frequent revision of many operational, engineering, and other basic assumptions, and the wider impact of support planning, requires that the support planning cycle be compressed in time. Such compression will help in the following ways. First, it is important in permitting the planners to achieve greater consistency in the different parts of the plan. Operational requirements, weapon characteristics, and other factors tend to change so rapidly during the development phase that the longer the planning cycle, the greater the opportunity for the various parts of the support plan to become incompatible and to be based on outmoded assumptions. Thus, one benefit of compressing the time required for support planning is the creation of more consistent plans by facilitating the use of joint planning from common assumptions and ground rules.

Another benefit of shorter planning cycles is that greater opportunity will be available to test alternatives. This flexibility is very important to assuring the development of a realistic and useful support plan able to deal with the uncertainty of conditions and capability that may arise when the weapon becomes operational. Thus, the development of a support planning technique which can deal in greater detail, proceed jointly, produce planning data more rapidly, and look at alternatives would contribute substantially to accelerating the operational availability of new weapons with economical use of resources.

The awareness of need for better support planning tools has always been present, and the benefits that improved support planning would provide have always been realized. However, the demands of the new environment are imposing tremendous stresses on the planning techniques. For example, under the present and foreseeable conditions, the need exists for techniques which can continually utilize experience based on components and subsystem testing in predicting weapon system performance and resource requirements since these data provide the only source for actual experience on failure rates, repair times, etc., at that stage in weapon system development when critical support planning decisions must be made.

Techniques for using these data, in combination with other planning data, to produce estimates of total resource requirements and operational capability are the essential requirement of the support planner. He has to ensure that the support implications based on these detailed development data are compatible with budget, program, reliability, and other grosser assumptions used by the higher echelon planning agencies in fitting this weapon system into the over-all support system. Thus, an important dynamic problem in support planning is the continual meshing of the over-all plans and their resource implications with the developing experience from testing, engineering studies, etc. On the one hand, this means projecting gross weapon system planning data to finer levels of detail to develop the detailed resource implications of the operational and support plans, and on the other, extrapolating detailed development data to grosser levels of aggregation to determine their implications for subsystem reliability, maintainability, etc. Either approach involves the risk of planning error, and essentially both must be used to mesh successfully planning and experience.

Simulation techniques offer promise in support planning. If a sufficiently realistic and valid simulated environment can be created at the time that the firm support policies and determinations of support policies are made, the planners can obtain a relatively rapid evaluation of these policies in terms of their implied weapon effectiveness and system cost. Even with the recognition of the inherent limitations of simulation techniques, nonetheless the environmental demands upon planning are such that the sound development of the simulation technique offers

the most promising opportunity for achieving a more consistent, accurate, and acceptable tie between resources and weapon effectiveness.

In the next section, we should like to describe a simulation experiment in which the detailed resource implications of general policies and planning assumptions were developed. The other part of this planning sequence—that of extrapolating from detailed development data—has not yet been done systematically. However, the first part of this sequence has been a significant enough accomplishment to deserve publication.

DESCRIPTION OF LABORATORY PROBLEM II

The first experiment of the RAND Logistics Systems Laboratory has been described elsewhere [1]. The second experiment was concerned primarily with the problem of support planning for new weapon systems. It took as its goal the development of a support system for the Intercontinental Ballistic Missile System of the 1963-65 time period. The approach was to simulate in detail those features of a support system that directly affect the operational effectiveness of the ICBM. We took these features to include the support organization and policies within the tactical command. The major organizational features were the number of missile complexes assigned to a single support center, the definition of the support functions performed at the missile launch complex, and those performed at the separate support center, and the management organization of the support center. The major support resources studied were maintenance personnel by type and level of skill, repair equipment by type, spare components by type, and transportation and other ground handling equipment by type. Under present Air Force plans, a 10-missile squadron has about 350 maintenance men, who use about 1000 pieces of repair equipment (with about 5000 major components in stock) and more than 200 pieces of ground-handling and transportation equipment. The cost of this support is about 20 million dollars per 10-missile squadron over a 5-year period, including initial investment.

The job then in LP-II was to specify the support features for this missile system 5 to 7 years in advance of its existence. To achieve this, we proposed to create a simulated environment of that time period, and to develop the preferred management system and support policies for living in that environment. By preferred we meant a system that incorporated the usual features of effectiveness, economy, and feasibility. We began first by creating an environment and testing in it the Air Force policies proposed for ICBM in the 1963-65 time period. We then used the results of this initial test as a bench mark for comparing variations in policy and organization. I shall not go into a detailed description of the simulation of the environment. This is described in [2].

The primary research problems posed in creating this simulated environment were, first, to what level of detail should the missile hardware be developed to permit an adequate interaction and evaluation of the support organization and policy questions described above. We found that we could treat these relationships satisfactorily by describing the missile, its ground support equipment, and critical facility elements in terms of what we called support units. Essentially, this meant subdividing the missile into "black box" or modular elements, and the facility and ground handling equipment into somewhat larger functional aggregations. This resulted in a final list of support units containing about 800 on the missile, about 250 on the launching and checkout equipment, 50 in ground handling equipment, and about 450 in the facilities, or about 1550 in all. The support unit list was organized to permit flexibility in modifying it as changes in subsystem specifications developed, as was the case during the study.

A second important research problem was to collect a vast amount of data about each support unit. These can be grouped under 4 broad headings: reliability, standard supply, maintenance, and operational maintenance status data. Approximately 100 numerical values had to be developed for each of the 1550 support units. Sophisticated techniques were called for in developing reliability and operational-maintenance status data. These are described in publication [2]. These techniques for converting operational and maintenance stresses placed on the missile to resulting missile down time and resource requirements are the essence of the simulated environment. The third important research problem was to prepare the experimental design that would define the kinds of simulation runs to undertake, their length, and the types of analyses to be made with the resulting data.

A run consisted of the following elements: the simulated environment created by means of computer programs, punch cards, and other abstractions; a group of Air Force personnel in the key management slots in the organization, and an information system useful to the Air Force managers for running their squadron, and to the laboratory staff for analysis. The Air Force personnel were given a week's training before they began the simulation. LP-II contained three major phases: The first phase was the run of the Air Force policies; the second phase introduced variations into the environment and the Air Force policies; and the third phase contained a system created from RAND research hypotheses determined from background research and earlier experience in the simulation.

As was stated earlier, the first phase for the Air Force policies provided the bench mark. This run lasted for two simulated weeks, with each fifteen-minute period providing an opportunity for a change in missile status, a demand for resources, or management action. Thus, there were 1344 fifteen-minute periods in the two simulated weeks. This was an ample length of time for the system to settle down and provide stable results.

The second phase began with a run lasting one simulated week with conditions varying from the first phase only by adding considerably to the operational requirements. The resulting impact on missile effectiveness and resource use was obtained. Here, too, the run length was sufficient to provide reliable results. The second variation in this phase, also lasting one simulated week, kept conditions as they were in the previous run, but reduced the assumed missile system reliability by 50 percent. Thus, the purpose in both the second and third runs was to vary the assumed environment to stress the support system. In the third variation of the second phase, also lasting a week, we varied the support policies by reducing the maintenance personnel of the squadron by about 25 percent, and the repair equipment by about 30 percent. Thus, the purpose of the last run was to obtain better support policies by using the earlier simulation experience and rational rules for allocating resources as a guide.

The third phase involved a major change in the support organization, support policies, and management system. Essentially, the notion was that by using the experience of the earlier phases as well as background research it appeared that a more efficient support organization and policies would be obtained if the number of missiles assigned to the support center were increased from 10 to 40. It seemed that if this were done, support costs could be reduced by 50 percent. We could define in advance desirable changes in the manning, equipping, spare components, and transportation equipment resources assigned for achieving this goal, but we were not certain of the impact of these changes on management activity and the management system in terms of organization, information requirements, and decision-making stresses. The purpose of the third phase was therefore to study intensively the management features of

this system, recognizing that management capability can substitute for resource requirements, and further that the creation of a management system may involve as long a lead time as the development of the weapon itself. Therefore, its characteristics and development should be an inherent part of the planning process.

From the first two phases, many minor but over-all significant improvements had developed in the simulation process, the simulated environment, and the application of rational resource policies. It was decided that for experimental control to run concurrently the 10-missile squadron with the 40-missile wing, to give them the same military job to do, assign resources to them in accordance with the same policy concepts, give each the best available management system, and then compare the two organizations as to operational effectiveness obtained and over-all system cost. This phase was run for almost three simulated weeks, during which tremendous operational and support stresses were placed on both organizations, stresses probably much greater than any that would be experienced in any comparable real-world period, but stresses that would test the flexibility, adaptability, and responsiveness of the management process.

This run required a total of 26 Air Force participants compared with the 6 for each of the first two phases, with 20 being assigned to the wing and 6 to the squadron. The 3 simulated weeks took 6 real weeks to cover. Tremendous amounts of data were developed during the run that are now being analyzed and will be reported elsewhere.

CONCLUSION

LP-II began in earnest about September 1958. Its runs were made in December 1959, and January 1959, for the first two phases, and June and July 1959, for the third phase. During this period a weapon system was subjected to more intensive and systematic research from a management and support system standpoint than had probably been done heretofore anywhere. This research has produced many specific findings of direct importance to the missile program, and these are now being reported to the Air Force. In addition, we believe that much progress has been made in systematically developing a support plan for a new weapon system, particularly one containing a radically different environment from the past. The technique used with the missile system can be applied to other important weapon systems. With the knowledge gained from LP-II in support concepts and policies, management systems and simulation techniques, the study for newer weapon systems can be done even more rapidly than was the case in LP-II. Against this, we must place the ability of the real world to relate operational experience to policy which is both unsystematic and requires several years to achieve to the limited degree it is done.

Also, the technique on an even broader base has tremendous applicability to the design of man-machine control systems, organizational and information research, and decision theory. In essence, in LP-II we created an environment which, while not as complex as the real world, is nevertheless sufficiently complex to challenge research workers in these fields to develop analytic techniques to describe the phenomena observed and to extrapolate the results from the simulation experiment to other conceivable environments.

REFERENCES

[1] Geisler, M.A., "A First Experiment in Logistics System Simulation," Naval Research Logistics Quarterly, March 1960.

[2] Haythorn, W.W., "Laboratory Simulation of a Ballistic Missile Squadron: A Description of LP-II, Phase 1.1," The RAND Corporation, Paper P-1656, February 17, 1959.

* * *

THE USE OF SIMULATION TO EVALUATE A MULTIECHELON, DYNAMIC INVENTORY MODEL

Andrew J. Clark

Planning Research Corporation
Los Angeles, California

Simulation is normally used to investigate a problem that is too complex for analytic solution. However, it is possible to have a solution algorithm that is itself so complex that simulation may profitably be used to study its behavior. A case in hand is a recently developed multiechelon inventory decision model which uses dynamic programming techniques. The use of simulation to evaluate this decision model is illustrated by a case study of a fictitious but realistic item. In particular, its use for comparative analyses of different maintenance policies is explained.

Additional uses of simulation in this context are mentioned. For example, the insertion into the simulation of complexities not included in the decision model provides a qualitative evaluation of the use of the decision model in an actual logistics system. Also, the ability of the decision model to cope with increasing information may be analyzed.

INTRODUCTION

The normal application of simulation methods is in the study of problems that are too complex for solution by analytic techniques. It is usually thought that if a solution algorithm can be constructed, then simulation becomes unnecessary in providing answers to the problems. However, there is a valid use of simulation in the presence of a solution algorithm; namely, to evaluate the consequences of actually applying the analytic solutions in the real-life environment. In general, such simulation models may be structured in two ways. First, the simulation may reflect identically the same portrait of reality as does the analytic model in order to evaluate the solutions of the analytic model relative to the problem it represents. Second, the simulation may reflect a richer environment than does the analytic model in order to evaluate the solutions obtained from an abstraction of the real problem when applied in the real problem environment.

The first type of simulation used in this context is usually not necessary, because the sheer construction of the analytic model provides a means for evaluating the consequences of applying the solutions under conditions assumed by the model. However, it sometimes happens that the analytic model and its solutions are so complex that even though its evaluation can be done analytically, it becomes easier and more economical to do it by simulation. A case in hand is a multiechelon, dynamic inventory model now under development at Planning Research Corporation.

Without explicitly describing the inventory model itself, the use of simulation in the context mentioned above is illustrated by three realistic but merely fictitious examples. Each example involves supply support for a missile-guidance gyro assembly used on missile-carrying submarines. Logistic support for the submarines is provided by submarine tenders, a repair depot, a supply depot, and a factory that produces the gyros. The repair and supply

depots and the factory must also support a personnel training program, a missile test program, and two shipyards involved in the outfitting and shakedown of new submarines.

To illustrate a variety of situations, the nature of the repair support is varied among the three examples. Case A considers that all reparable gyro failures are repaired at the repair depot, while Case B assumes that 60 percent of reparable failures are repaired at the using activities, the remainder going to the depot for repair. Case C considers that gyro as a nonreparable item.

The remainder of the paper is divided into three sections. The first deals with a detailed description of the hypothetical item, logistic support structure, pertinent parameters, and computation of inventory levels. The technique of simulation, a method for making inventory decisions with pseudohistorical data to evaluate the policies, is described in the next section. The results of the three simulations and an analysis thereof are presented under the third major heading. This section portrays particularly the type of analyses that are easy to obtain through simulation, but difficult to establish analytically. The appendix contains a general equation for computing the expected regeneration time for the three examples given in the text.

EXAMPLE PROBLEM

Description of Item

The item considered here is a missile-guidance gyro assembly designed for submarine-based missiles. Its unit cost is $2000 delivered to the supply depot. The gyro is assumed to have a two and one-half year mean service life with small premature failure characteristics as shown in Figure 1. Each installed gyro is assumed to be replaced with a successor model two and one-half years after initial installation.

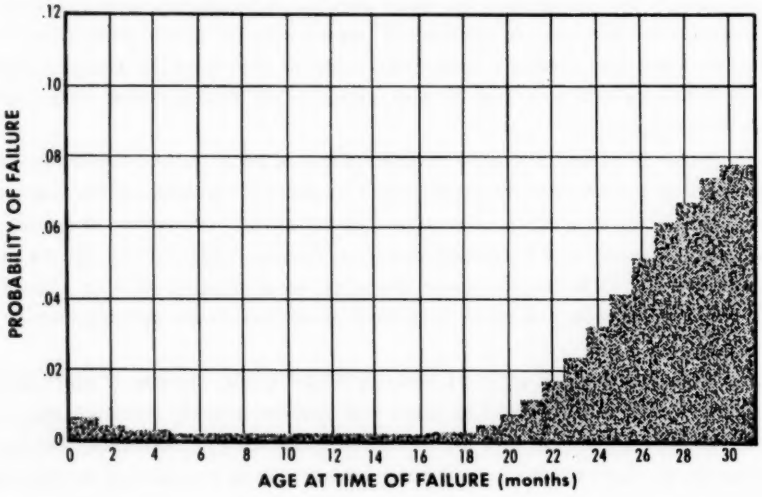


Figure 1 - Probability of failure as a function of gyro age

As mentioned before, gyro reparability varies among the three examples—the variability is shown in Table 1. Case A considers that 10 percent of the failures result in condemnations and the other 90 percent are sufficiently difficult to repair that depot repair is required. Case B also considers that 10 percent of the failures result in condemnations, but not all of the other 90 percent of the failures result in difficult repairs. It is assumed that 60 percent of the repairs are simple and can be accomplished at the using activity, while the other 30 percent are difficult and must be returned to the repair depot. In Case C the gyro is considered nonreparable (i.e., 100 percent of the failures result in condemnations).

TABLE 1

Gyro Reparability

Case	Percentage of Failures that are ...		
	Repaired at Using Activities (%)	Repaired at Depot (%)	Condemned(%)
A	--	90	10
B	60	30	10
C	--	--	100

Logistic Support Structure

The geography of the logistic support system, including pertinent parameters and the flow of spare gyros, is shown in Figure 2. Details concerning the various activities are discussed below.

Using Activities. There are eight activities that use missile-guidance gyros in these hypothetical problems: four submarine tenders, two shipyards, a test program, and a training program.

In reality, tenders are not using activities. That is, demands for items do not normally originate aboard tenders, but aboard the submarines supported by the tenders. For purposes of these examples, though, demands are assumed to originate aboard the tenders.[1] This is because the model, as presently programmed, will not accept more than one activity laterally except at the lowest level in the structure, viz., the using activity level.

Each tender is assumed to support six submarines and each submarine carries 16 applications of the missile gyro. Tenders A and B and 12 submarines are located in the Atlantic Ocean, and Tenders C and D and 12 submarines are located in the Pacific Ocean.

The shipyards generate demands during the outfitting and shakedown of new submarines. Perhaps a gyro is damaged during transportation or installation or maybe it should not have passed final inspection at the factory. One shipyard is located at Port Charleston, South Carolina, for the outfitting and shakedown of Atlantic-based submarines, and the other is located at Mare Island, California, for support of Pacific-based submarines.

Support System. The repair depot is located at Port Charleston, South Carolina, physically attached to the supply depot. All failures too difficult to repair at the using activities are sent here for repair. These returns, for purposes of clarity, are not shown in Figure 2.

[1] This assumption is reasonable because items which fail aboard a submarine are replaced out of tender stocks at the time the submarine returns to the tender.

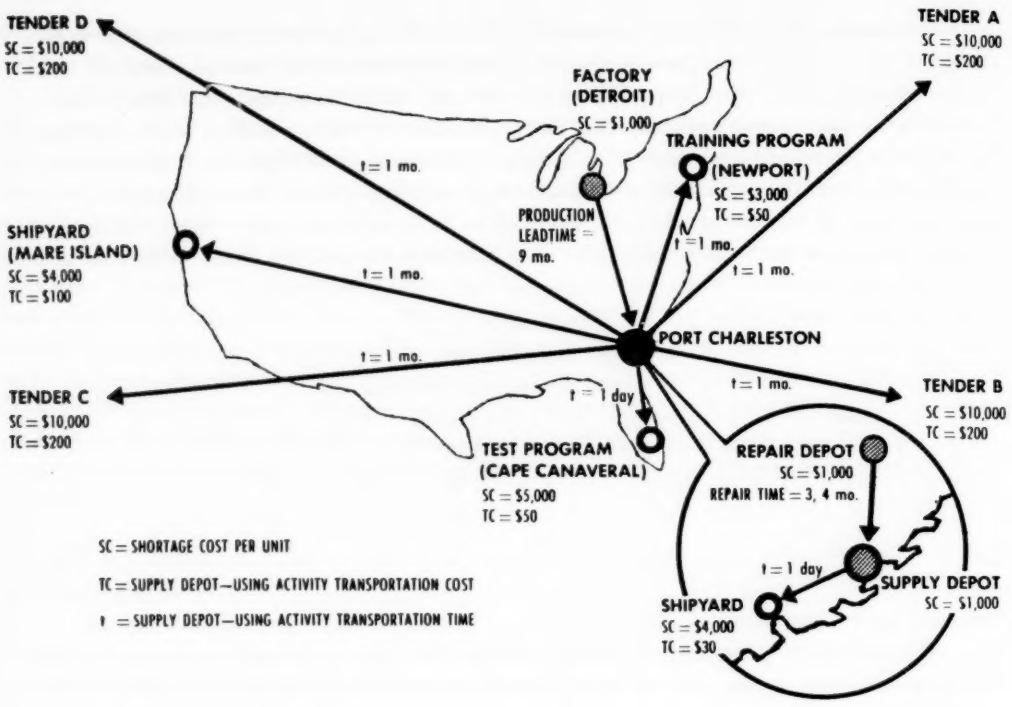


Figure 2 - Logistic support system

For the first two years of the five-year program the repair time is three months per item and the repair cost is $300. During the last three years of the program a more extensive repair and overhaul requirement is assumed due to the increased age of the item at time of failure. Accordingly, during this portion of the program, the repair time and cost are taken as four months and $1000, respectively.

The factory keeps its gyro production facilities set up during the first two years of the program and then tears them down. During the first two years the fixed cost of reordering is $500, but after the set-ups have been torn down the fixed reorder cost is $50, 000.

The unit cost of the gyro builds up incrementally during the nine-month production lead time as indicated in Figure 3. During the first three months of the lead time very little cost is incurred, the cost mainly stemming from the paperwork cost of ordering material and scheduling production. From the fourth to sixth months, 75 percent of the cost of the gyro is incurred; fabrication takes place during these months. Testing and inspection take place during the last three months of the lead time and, accordingly, relatively little additional cost is incurred.

The holding cost rate, also shown in Figure 3, is dependent on the space the gyro occupies on the factory floor and the obsolescence risk. The holding cost rate is constant at

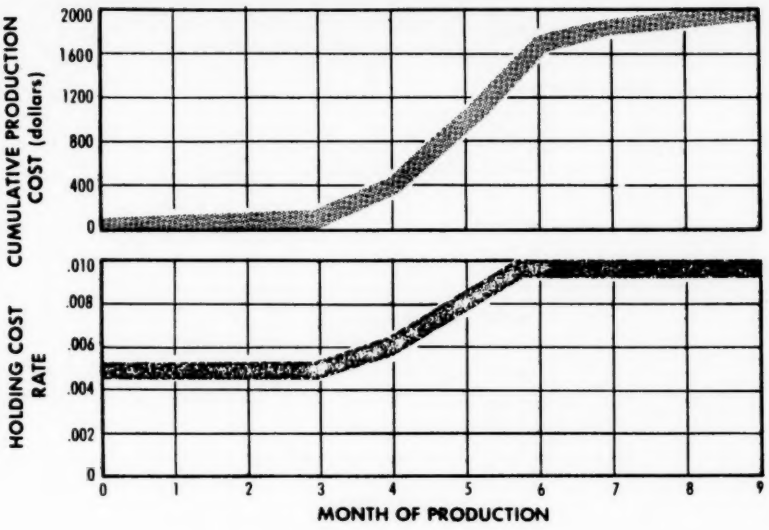


Figure 3 - Factory parameters

a low value for the first three months of the lead time, reflecting the compactness and relatively high surplus value of raw material. During the fabrication period the gyros begin to occupy more space and decrease in surplus value until the holding cost rate finally doubles after six months. The holding cost remains constant at the doubled value during the testing and inspection period. The interest rate, for purposes of discounting future costs, is 2 percent per month.

Dynamics

The time phasing of the missile gyro at using activities appears in Figure 4. The program is seen to be five years long. This is not meant to imply that the submarines are retired after five years, but rather that a successor has gradually replaced the older gyros as the submarines have been overhauled at the end of two and one-half years of service.

Based on Figure 4, a graph showing the parts population of installed gyros as a function of time was constructed and is portrayed in Figure 5. The parts population curve for Tender A indicates three distinct periods: a phase-in, a steady state, and a phase-out. An additional submarine becomes operational every two months during the phase-in and each of the six submarines remains operational until time for overhaul at two and one-half years of service. Since the gyros are replaced with successor models during overhaul, the installed gyro population for Tender A diminishes at the overhaul rate. The installed gyro populations for Tenders B, C, and D are identical to that of Tender A; the others are merely displaced in time.

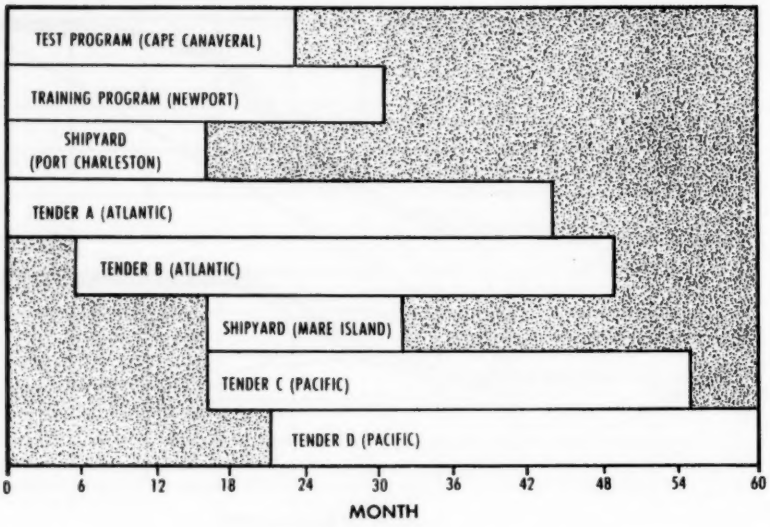


Figure 4 - Time phasing of missile gyro at using activities

The test program is scheduled to launch 300 missiles, but the supply system does not assume support until 125 missiles have been fired. The remaining 175 missiles are launched over a two-year period. The graph in Figure 5 shows the number of missiles launched during a specified month.

Six missiles are at the disposal of the personnel training program. The missiles are used in launching drills, but are not fired. The six missiles contain six gyros which are replaced with successor models after two and one-half years (the time when the last submarine attached to Tender D is phased in) since further training with the earlier model gyro is unnecessary.

The shipyards outfit and shake down one submarine at a time, on the average. The curves in Figure 5, therefore, show a constant 16 missiles in use at both shipyards. The Port Charleston shipyard ceases to need supply support when the last Atlantic-based submarine is phased in at 16 months. At two and one-half years, when the last Pacific-based submarine is phased in, the Mare Island shipyard ceases to need supply support.

Demand Estimates

Estimates of expected monthly demand for gyros at the using activities are shown in Figure 6. The number of demands during any month is assumed to be Poisson distributed.

Mean demands for spare gyros from the submarine tenders are estimated with the information in Figures 1 and 5. It is assumed that when a gyro fails aboard a submarine it is replaced with one having the same age. Thus, Figure 1 gives the probability of gyro failure regardless of the number of previous replacements. [2]

[2] Mean demands could have been estimated under the assumption that all replacements involve new gyros, but the estimation procedure would have been complex and the estimates would not be significantly different.

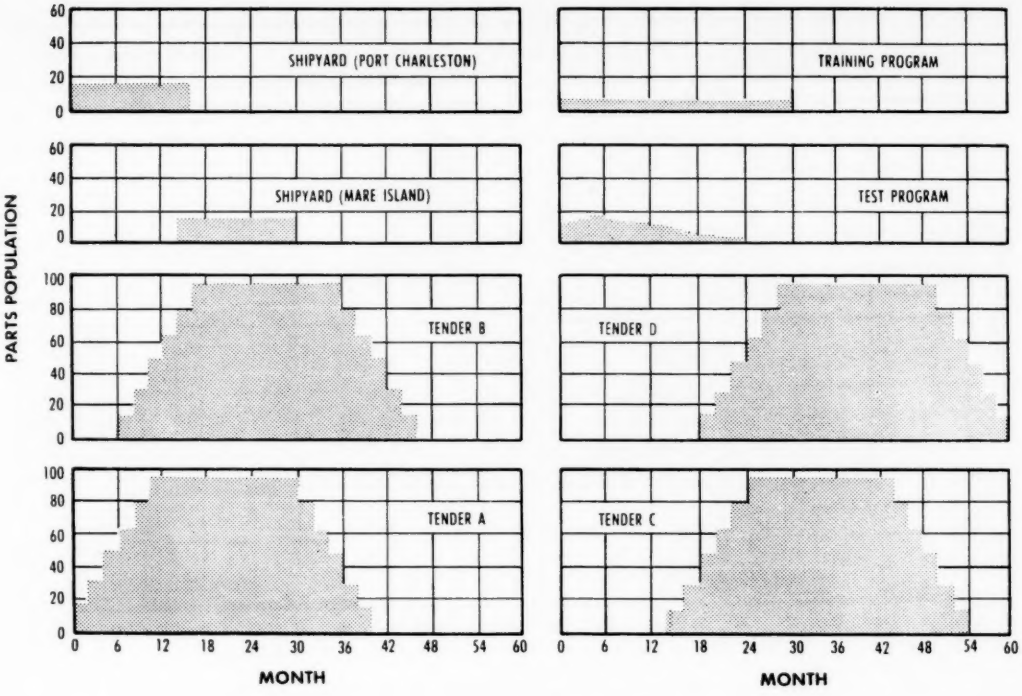


Figure 5 - Parts population of installed missile gyros

Since there are 16 gyros aboard each submarine, each ordinate in Figure 1 is multiplied by 16 to obtain expected monthly demand per submarine. When the expected monthly demands for six submarines are arranged in the proper columnar form (bearing correspondence to the phase-in) and added, the result is expected monthly demand at a tender.

The probability density for gyro failure is not used in estimating gyro demands at ancillary using activities. The expected demands are taken as proportional to the parts population. For instance, to generate expected demands for the test program it was assumed that approximately two gyros would be demanded for every five missile launchings. The training program puts six missiles through intermittent use for 30 months; assuming that each missile gyro fails once, the expected monthly demand is $6 \div 30 = 0.2$. For the shipyards, it was assumed that an average of one gyro would be demanded per outfitted submarine.

The summation of using activity expected monthly demands appears in Figure 7. This is the total expected demand that must be satisfied by the supply system.

Computation of Levels

The characteristics of the item and its logistic support system discussed above were transformed into computing machine inputs for the calculation of inventory decision levels.

Two inventory decision levels are determined for each activity during each month: a reorder point and a stock control level. If, at the beginning of a month, an activity is at or below the reorder point, it orders a sufficient number of gyros to bring it up to the stock control level (provided the required amount is available from the next highest echelon).

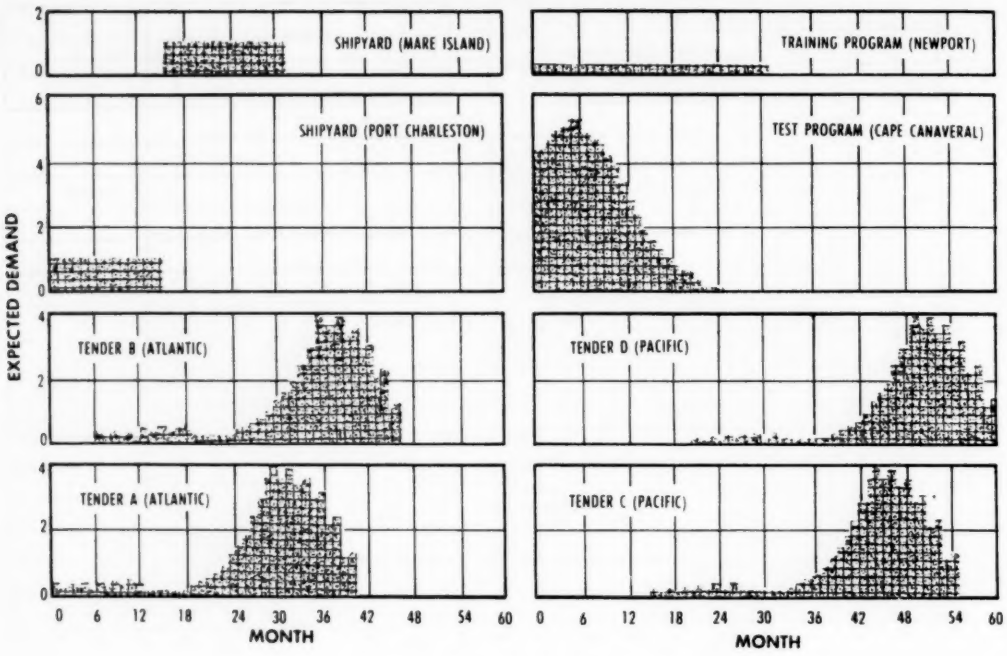


Figure 6 - Expected monthly demand at the using activities

For all activities except the highest factory echelon the reorder point is always one unit less than the stock control level. At the highest factory echelon the reorder point can be more than one unit less than the stock control level. The latter reorder point indicates the need for procurement if total system stock is at or below this level.

SIMULATION

The inventory decision levels yield little information about the expected performance of the supply system or its associated costs. To gain insight into performance and cost it is useful to perform a "simulation." Simulation involves allowing randomly-determined using-activity demands to act on the supply system and making inventory decisions in accordance with the computed decision levels.

The using-activity random monthly demands are shown in Figure 8 superimposed on the expected monthly demands. Superimposition illustrates the effect of uncertainty. With the information in Figure 8 as a base, reparability characteristics of the demand are determined. In Case A, 10 percent of the demands are supposed to result in condemnations. For each demand a random number was drawn from a population with 10 percent of the numbers predesignated as condemnations. The demand was then designated as either depot reparable or condemned. A comparable procedure was used in selecting depot reparables, using activity reparables, and condemnations for Case B.

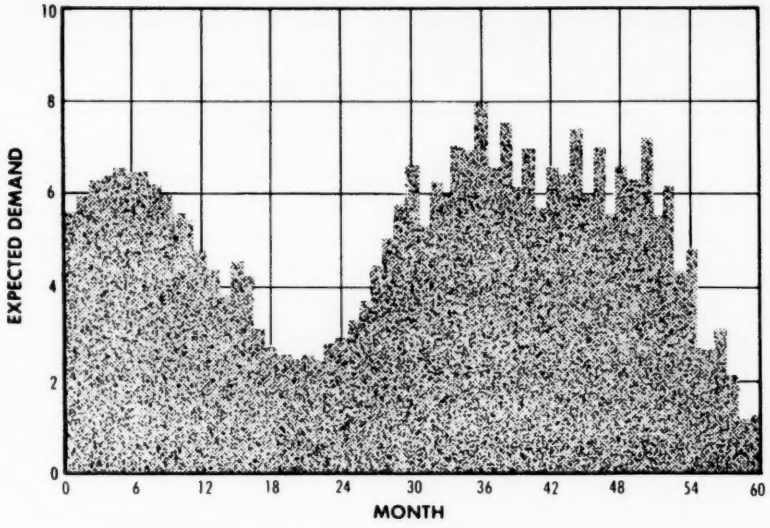


Figure 7 - Totals of using-activity expected monthly demands

Care should be exercised in interpreting the simulation results presented in the next section because they are based on only one set of random demands. The results should not be considered as an expectation, but rather as an example of what would happen given this particular set of demands with the system operating under the policies computed from the input parameters. More valid inference could be drawn from the averaged results of several such simulations, each using an independent set of random demands.

The simulation is performed entirely within the structure of the model with one exception. The model assumes, for ease of computation, that at the beginning of each month stock available at the depot is distributed to using activities so as to equalize protection against stockouts (weighted by shortage cost) or, when there is an insufficient amount at the depot, stocks at using activities are transshipped to equalize protection. In reality, however, such transshipment frequently would be impractical in view of the world-wide logistic support system used in the examples, e.g., an Atlantic-based tender would not normally send a one-unit stock overage to a Pacific-based tender. Transshipment between using activities was therefore not allowed in the simulation, although the inventory decision levels were computed as if such transshipment were allowed. Consequently, the levels give somewhat less protection from shortages in the simulations since levels are lower under a system of transshipment.[3]

[3] It should be noted, however, that no shortages were experienced in any of the three simulations. The main reasons for this were that relatively high shortage costs were assumed and that only 296 random demands were generated compared with an expected 302. It should also be noted that in two of the cases (depot repair and total condemnation), the model assumption of transshipment has little import since almost always a sufficient amount is available at the depot to equalize protection at the using activities when distributed.

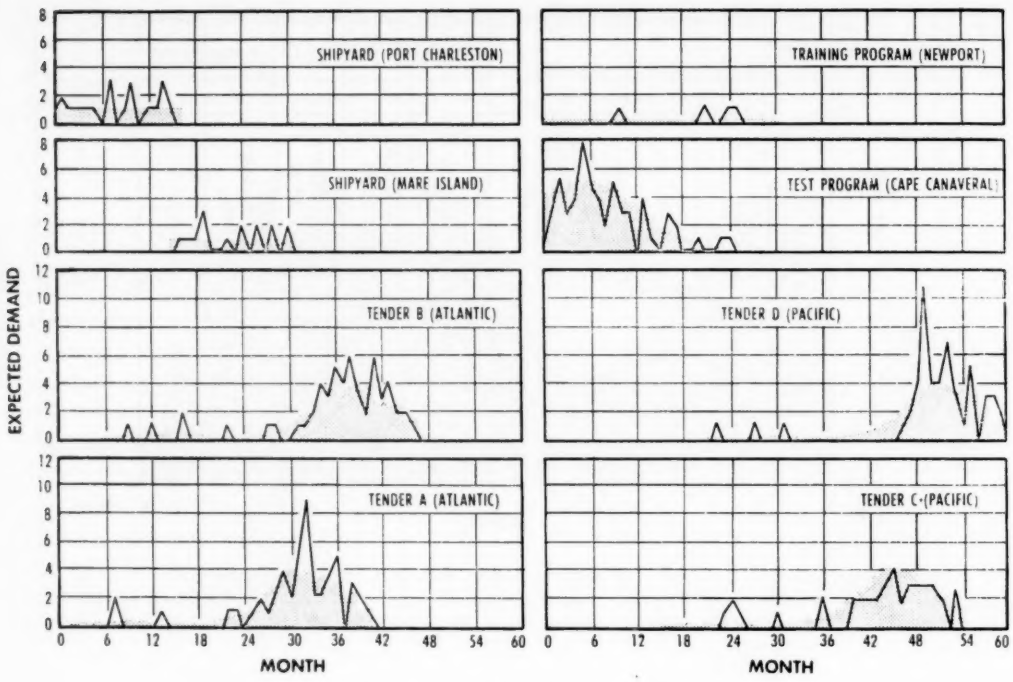


Figure 8 - Using-activity simulated and expected monthly demand

ANALYSIS OF SIMULATION RESULTS

Analysis of the simulation results is divided into two subsections: one concerning the system stock distribution as a function of time and the other concerning the manner in which funds are spent on supply support.

Stock Distribution

Figure 9 depicts the stock distribution as a function of time for all three examples. At the bottom of the figure, for comparision purposes, appears the random monthly system-demand.

All three examples exhibit similar factory ordering practices. Since the production lead time is 9 months, the first factory order occurs 9 months before the beginning of the program. Orders are placed during most of the months before the beginning of the program in each example, and little factory inventory is retained during this period, most of the items being processed to completion so that initial demands can be satisfied.

A life-of-type purchase is made in month 24 in all three examples. The inventory decision levels actually force this buy as a protection against reordering after month 24, when the fixed reorder cost is $50,000 instead of $500.

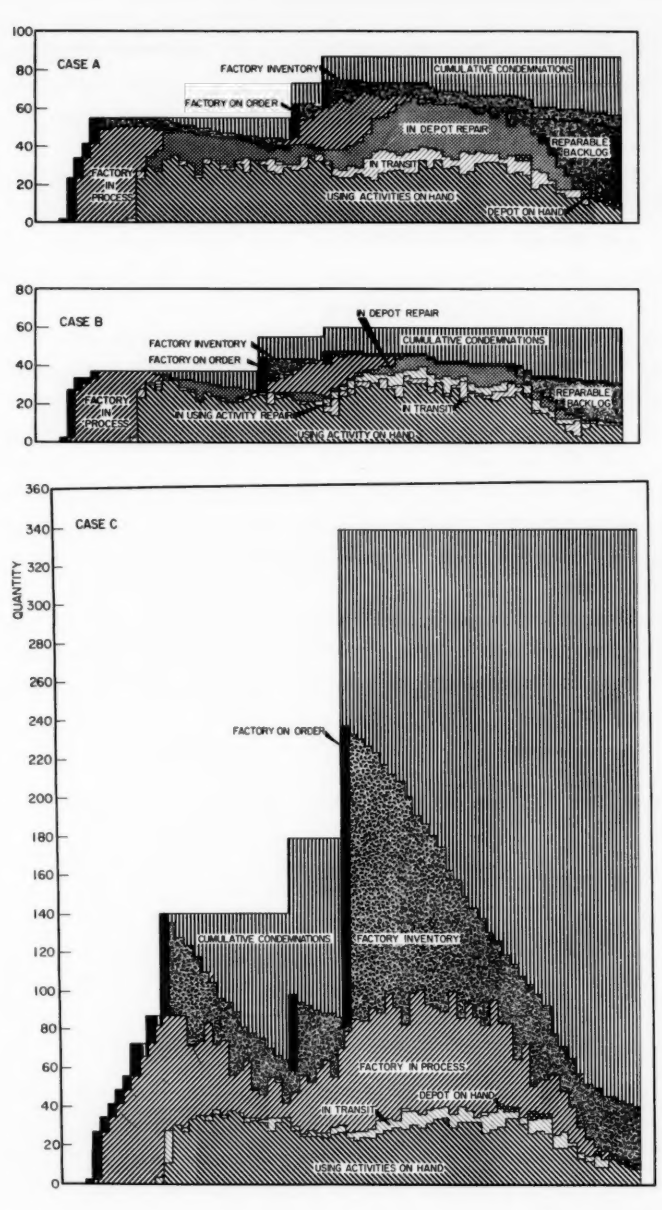


Figure 9 - System stock distribution as a function of time

In each example, one reorder is made between the beginning of the program and month 24. These intermediate reorders occur approximately one and one-half years into the program. Evidently, holding costs prohibit buying a sufficient number of items before the beginning of the program to last until the life-of-type buy in month 24.

Reparable backlogs build up in the two reparable cases during periods of decreasing demand. This is because the system needs many spare gyros for protection from shortages during periods of peak demand, but the protection is no longer needed when the demand starts to decrease. Thus, it is not economical to repair gyros that fail during the peak demand period, so the failed gyros accumulate in a backlog.

The supply depot rarely keeps items on hand; it usually passes them on immediately to the using activities.[4] The supply depot keeps items on hand only when a using activity phases out with spare gyros on hand and the rest of the using activities are stocked up to level.

It is useful at this point to define a measurable index of the mode of replenishment. The index is the "expected regeneration time," and it refers to the expected time (from the time of a failure) required by the mode(s) of replenishment to reproduce a serviceable gyro and deliver it to a using activity. This is not to be confused with the time to replace a failed gyro, which might be either one month or zero, depending on the supply depot-to-using-activity transportation time.

Details of the expected regeneration time computation are irrelevant to this analysis, but the results are given in Table 2. The three cases have significantly different expected regeneration times; these differences will be used to explain some of the phenomena exhibited by the simulation results.

TABLE 2
Expected Regeneration Times

Case	Expected Regeneration Time, Months
A	4.91
B	3.15
C	9.73

The using activity on hand plus in-transit quantities for the three examples are shown in Figure 10. It is seen that these quantities are essentially independent of the expected regeneration times. This is because the using activities exhibit the same cost and demand patterns in the three cases and therefore create identical needs that must be satisfied, regardless of the mode of replenishment.

The quantity in the replenishment pipeline (i.e., in repair plus factory in process plus factory on order) is very dependent on the expected regeneration time as shown in Figure 11. As the expected regeneration time increases, the number of units in the pipeline increases, following the principle that more items are needed to fill a longer pipeline.

[4] This occurs because the savings in expected shortage costs afforded by having the item on hand at a using activity are usually large enough to pay the transportation cost from the depot. Furthermore, since the reorder quantity is one, a using activity falls below its level whenever a demand occurs.

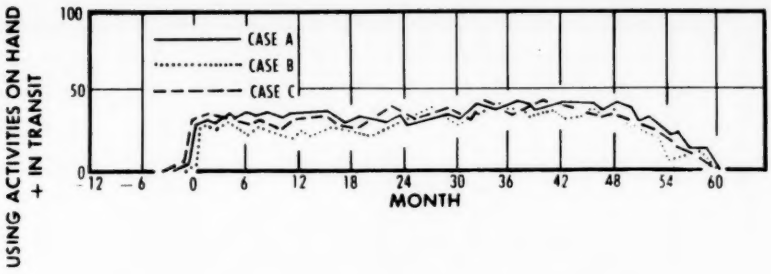


Figure 10 - Using activities on hand plus in-transit as a function of time

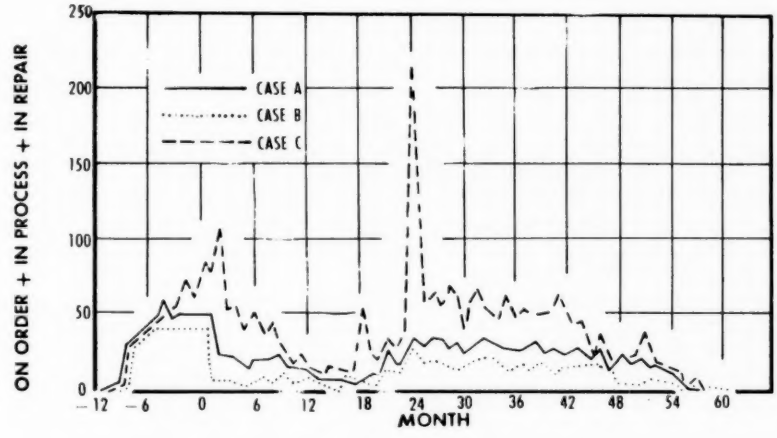


Figure 11 - Replenishment-pipeline quantity as a function of time

The level of factory inventory (raw material and partially finished goods) is also highly dependent on expected regeneration time as indicated in Figure 12. Such inventory is kept as protection against shortages and reorders, and is necessarily proportional to the protection needed. The protection needed varies directly with the demand expected during a lead time and this, of course, is the essence of the expected regeneration time calculation.

Supply Expenditure

The supply expenditure as a function of time for the three examples appears in Figure 13. For reference purposes, the random system monthly demands appear at the bottom of the figure. The value of such information to the budget planner is readily apparent; also such information is invaluable to the economist investigating problems like repair vs discard.

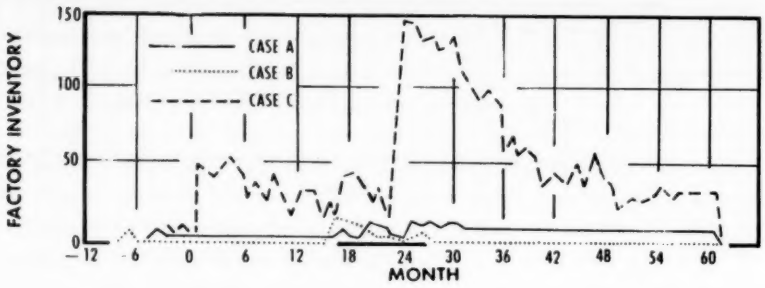


Figure 12 - Factory inventory as a function of time

The information given in Figure 13 also appears in summary form in Table 3. Discussion will center around the table.

TABLE 3

Supply Expenditure (Thousands of Dollars)

Type of Cost	Case		
	A	B	C
Reorder Cost	$ 4	$ 3	$ 6
Holding Cost	52	25	55
Process Cost	163	116	616
Depot Repair Cost	166	51	--
Using Activity Repair Cost	--	28	--
Transportation Cost	50	23	49
Total Cost	$435	$256	$726

Reorder cost for the three cases is seen in Table 3 to be relatively constant compared with the fluctuations exhibited by other types of costs. Holding costs are nearly the same for Cases A and C. At first, this seems counter to intuition since 86 items were ordered in Case A and 336 in Case C; it would seem that Case C should exhibit higher holding costs. Actually, many of the 336 items in Case C remain in either the raw material state or the partially finished state much of the time, and the holding costs in these states are small compared with the holding cost for a completed item. Completed gyro inventories are nearly the same in Cases A and C and are less in Case B, and as such give rise to the holding cost pattern exhibited in Table 3.

Process costs in Case C far exceed the process costs in Cases A and B. This is to be expected since, outside of supplying the gyros that cycle in repair, the factory satisfies only 10 percent of the demands in Cases A and B. In Case C, the factory satisfies all the demands.

Repair costs in Case A are $166,000 and in Case B, $79,000. This difference is merely due to the difference in repair cost assumed in the two situations. Case A repairs

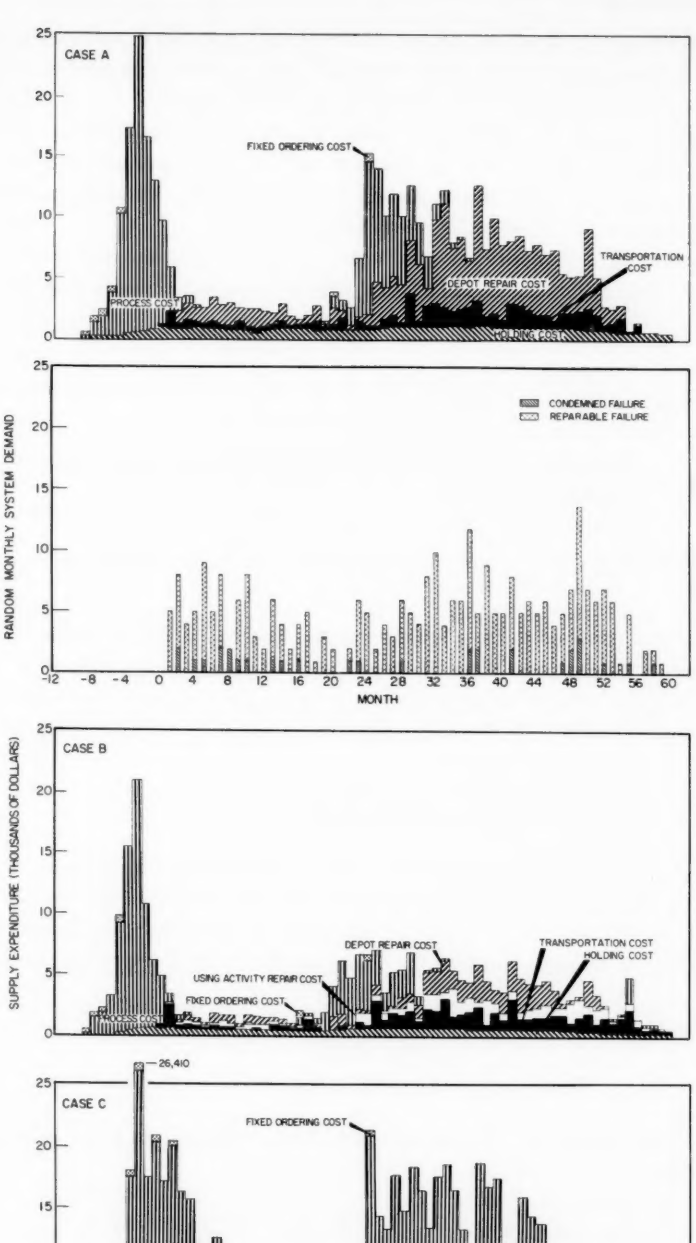


Figure 13 - Supply expenditure as a function of time

are expensive ($300 for the first two years, $1000 for the last three years); whereas in Case B, only one-third of the repairs incur this high cost, the other two-thirds incurring the lower activity repair cost (from $30 to $200, depending on the activity).[5]

Depot-using activity transportation costs are practically identical in Cases A and C and are less in Case B. This is because Cases A and C have 90 percent and 100 percent of the demands, respectively, satisfied by higher echelon sources, while in Case B only 40 percent of the demands are satisfied by higher echelon sources.

It is interesting to compare the total costs in Table 3 with the expected regeneration times in Table 2. The following ratios of total cost to expected regeneration time are obtained:

$$\text{Case A: } \frac{\$435,000}{4.91 \text{ mo.}} = \$88,600 \text{ per month of expected regeneration time.}$$

$$\text{Case B: } \frac{\$256,000}{3.15 \text{ mo.}} = \$81,400 \text{ per month of expected regeneration time.}$$

$$\text{Case C: } \frac{\$726,000}{9.73 \text{ mo.}} = \$74,700 \text{ per month of expected regeneration time.}$$

Little significance is attached to the absolute values, but the clustering tendency of the three ratios clearly indicates the possibility of using expected regeneration time as a linear cost predictor. Of course, a Monte Carlo analysis, rather than a one-draw simulation, would have to be performed to estimate the efficiency of this seemingly linear correspondence.

REFERENCES

[1] Clark, A. J., "A Dynamic, Single-Item, Multi-Echelon Inventory Model," RM-2297, Santa Monica, California, The RAND Corporation, December 1958.

[2] Clark, A. J., and Scarf, H., "Optimal Policies for a Multi-Echelon Inventory Problem," PRC R-5113, Los Angeles, California, Planning Research Corporation, July 1959.

APPENDIX

Computation of Expected Regeneration Time

For the three examples in the body of this paper a general equation for expected regeneration time is:

$$\begin{pmatrix}\text{expected}\\ \text{regeneration}\\ \text{time,}\\ \text{months}\end{pmatrix} = \begin{pmatrix}\text{fraction of ex-}\\ \text{pected failures}\\ \text{that are condem-}\\ \text{nations}\end{pmatrix} \begin{pmatrix}9\\ \text{months}\end{pmatrix}$$

[5] The structure of the model requires that the charge for a using activity repair be equal to the depot-using activity transportation cost.

$$+ \begin{pmatrix}\text{fraction of}\\ \text{expected}\\ \text{failures}\\ \text{that are}\\ \text{depot}\\ \text{repairs}\end{pmatrix} \left[\begin{pmatrix}\text{fraction of}\\ \text{expected}\\ \text{failures}\\ \text{that are}\\ \text{3-month}\\ \text{repairs}\end{pmatrix} \begin{pmatrix}3\\ \text{months}\end{pmatrix} + \begin{pmatrix}\text{fraction of}\\ \text{expected}\\ \text{failures}\\ \text{that are}\\ \text{4-month}\\ \text{repairs}\end{pmatrix} \begin{pmatrix}4\\ \text{months}\end{pmatrix} \right]$$

$$+ \left[\begin{pmatrix}\text{fraction of}\\ \text{expected}\\ \text{failures}\\ \text{that are}\\ \text{using}\\ \text{activity}\\ \text{repairs}\end{pmatrix} + 1 \right] \left[\begin{pmatrix}\text{fraction of}\\ \text{expected}\\ \text{failures}\\ \text{occurring}\\ \text{at transpor-}\\ \text{tation time}\\ = 1 \text{ month}\\ \text{using}\\ \text{activities}\end{pmatrix} \begin{pmatrix}1\\ \text{month}\end{pmatrix} + \begin{pmatrix}\text{fraction of}\\ \text{expected}\\ \text{failures}\\ \text{occurring at}\\ \text{transporta-}\\ \text{tion time =}\\ 0 \text{ using}\\ \text{activities}\end{pmatrix} (0) \right]$$

For Case A, the specific equation is:

$$\begin{pmatrix}\text{expected}\\ \text{regeneration}\\ \text{time}\end{pmatrix} = (.1)(9) + (.9)\left[\left(\frac{109}{303}\right)(3) + \left(\frac{194}{303}\right)(4)\right]$$

$$+ (0 + 1)\left[\left(\frac{220}{303}\right)(1) + \left(\frac{83}{303}\right)(0)\right] = 4.91 \text{ months.}$$

For Case B, the specific equation is:

$$\begin{pmatrix}\text{expected}\\ \text{regeneration}\\ \text{time}\end{pmatrix} = (.1)(9) + (.3)\left[\left(\frac{109}{303}\right)(3) + \left(\frac{194}{303}\right)(4)\right]$$

$$+ (.6 + 1)\left[\left(\frac{220}{303}\right)(1) + \left(\frac{83}{303}\right)(0)\right] = 3.15 \text{ months.}$$

And for Case C, the specific equation is:

$$\begin{pmatrix}\text{expected}\\ \text{regeneration}\\ \text{time}\end{pmatrix} = (1.0)(9) + (0)\left[(0)(3) + (0)(4)\right]$$

$$+ (0 + 1)\left[\left(\frac{220}{303}\right)(1) + \left(\frac{83}{303}\right)(0)\right] = 9.73 \text{ months.}$$

* * *

ON-LINE MACRO-SIMULATION IN SYSTEMS FOR LOGISTICS DECISION MAKING

Francis Dresch

Stanford Research Institute

Macro-simulation has three types of application:

1. Computer experimentation with simplified models of complex systems to explore the systems' sensitivity to variation in environmental parameters or changes in controllable decision variables. This use of macro-simulation facilitates analysis of the realism of the various models as well as determination of optimal strategies.
2. Pilot testing of complex decision systems against a detailed quantitative description of the real operating environment as observed in the past.
3. Continuing on-line simulation of a system over a future planning period to determine optimal levels or settings for current decision variables or to provide measures of effectiveness for feedback and control.

While the first two applications are most familiar and are most frequent, the third application may have increasing importance in the future. The reasons for this will be discussed with illustrations from work at Stanford Research Institute on stock level policy and on air logistics. All three applications of macro-simulation techniques have proved important in this work.

This panel is concerned with macro-simulation. As I understand the distinction that we are making between micro- and macro-simulation, macro-simulation refers to any attempt to simulate a full-scale logistics operation while micro-simulation is more closely synonymous to the original use of the expression Monte Carlo technique, that is, simulation of one element of a complex problem or the use of random-process analogues to evaluate stochastically a single mathematical equation.

With this definition in mind, the uses for macro-simulation fall in one or more of three categories:

1. Experimentation
2. Pilot test
3. Continuing (on-line) control

Any macro-simulation model must include a mathematical description of the environment concerned, a mathematical formulation of alternative decision rules whose effects are to be simulated, and the specification of individual decision variables corresponding to possible decisions. Experimentation may concern alternative descriptions of the environment, alternative specification of the parameters in such an environmental model, or alternative decision rules and decisions. The exploration of the effects of changes in environmental parameters or relations is the essence of sensitivity analysis and provides quantitative information

as to the realism of the simplified environmental models to be used for exploring the effects of alternative strategies.

Experimentation with alternative strategies provides measures of the relative effectiveness of such strategies and permits an experimental approach to optimization. Optimization and sensitivity analysis together supply the requirements for simulation as an experimental tool.

Pilot testing of a system based on a selected strategy can be distinguished from experimental simulation primarily in that the model used for the environment in the pilot test may be more realistic, more detailed, and even structurally quite different from that used for experimentation and optimization. The differences arise because naive optimism, theory, mathematical manipulation, or sensitivity analysis may suggest drastic simplification of the experimental model that will help economize on the experimental effort required for optimization. Since it is the validity of these simplifications that must be verified in the pilot tests, the models used for test must be more realistic. Moreover, pilot tests should reproduce actual operating situations as generally familiar to operating people and provide more convincing validation of strategies than would the more abstract formulations most effective for experimentation.

The use of simulation in continuing (on-line) control is rarer than the more familiar uses in experimentation and test. The need for simulation of this type arises whenever the best current setting of decision parameters must be obtained from a comparative evaluation of their probable effects on an objective function over a future period. The simulation measures the relative aggregate effectiveness of currently available decision alternatives. Simulation may also be necessary to complete a feedback loop desired for monitoring an on-line system. This is necessary or desirable when feasible operating statistics do not permit adequate estimates of values being achieved for an objective function.

These and other applications of simulation can best be illustrated by brief discussion of two examples employing simulation in all these forms. The first example concerns our studies at Stanford Research Institute for the U.S. Navy Bureau of Supplies and Accounts on safety level policy for replenishable demand. The second example concerns an airlift study conducted for the Air Force.

EXAMPLE FROM SAFETY LEVEL STUDY

A multiechelon supply system such as that managed by the Navy's complex of Supply Demand Control Points must contend in the aggregate with over a million stock items, each stored at perhaps 20 or 30 different points. Its decisions with respect to these items concern placement of procurement orders, advance allocation of deliveries to stocking points, reallocation of orders not yet shipped, and redistribution of material from one stocking point to another.

In early experimentation with this problem, the description of the environment was formulated in several different ways all greatly simplified. For some purposes, it sufficed to experiment with a single item and a single stocking point—virtually micro-simulation. For experimentation with redistribution, the case of a single item stocked at several places was examined. For system-wide procurement, analysis was made of allocations of available funds over a variety of items. Throughout such studies demands were generated by the Monte Carlo process from various assumed distributions or, where lead time distributions were of particular interest, from formulas for expected demand. Similarly, lead times were generated either by Monte Carlo processes in the computer or, when attention was directed toward experimentation with demand distributions, by mathematical formulas.

The environmental models used have assumed various forms, and this will happen in general. Sensitivity analysis applied to all such variants is a necessary step in narrowing the alternatives to one or two environmental models.

Parallel to the effort to describe the environment must run an effort to settle on one or more objective functions, that is, one or more measures of effectiveness of the strategies to be selected for careful comparison and evaluation. In the example being discussed a measure described as total essentiality-weighted unit-weeks of shortage counted at all stocking points was adopted as an objective function to be minimized within the limits of available total funds. The phrase used to describe the objective function (which we will abbreviate to TEWUWS) requires some explanation. Perhaps a word-by-word discussion will suffice:

> Total refers to summation over all items and all stocking points and over some specified time period.
>
> Essentiality-weighted refers to a weighting of shortages according to assigned importance per unit of different items, or for weighting shortages at different stocking points.
>
> Unit-weeks refers to the fact that the shortages are measured both by the extent of shortage (the number of units short) and the duration of shortage (the number of weeks a requisition is delayed).
>
> Shortage is measured whenever activity stock levels drop to zero or to a mobilization reserve. System-wide shortages are sometimes referred to and occur when total stock in the system falls to zero. This has meaning only when aggregative models for the whole system are considered, ignoring redistribution possibilities.

Without attempting here to defend this particular choice of objective function (alternatives have been proposed and other alternatives are being investigated), the sensitivity testing consisted in experimental investigation of the effects of changes in environmental parameters on the TEWUWS. The general form of the decision rules investigated were an X, Q (order point, economic order quantity) policy applied either to individual activities or to system-wide procurement. The optimization principle followed was to attempt to minimize the TEWUWS for specified total costs in inventory holding charges, set-up and order costs, freight charges, etc., based on cost factors given as parameters in the environmental model. Rather than searching for the minimum TEWUWS, however, a simpler approach was employed which consisted of assuming a cost equivalent for a contribution of one unit to TEWUWS and then attempting to minimize total costs for each item including imputed shortage costs. This cost equivalent of a shortage was assigned the Greek letter λ because of its mathematical role as a Lagrangian multiplier. The variable λ could have been regarded as an environmental parameter to be determined by evaluation of the cost or average penalty incurred per shortage unit. In practice it was treated as a decision variable to be adjusted to equate total costs to a desired budget. Curves were obtained relating λ to total costs under alternative decision rules from which a suitable λ could be selected to achieve a suitable budget level.

The formula for total cost per unit time may be written as

$$ac(X+Q/2) + A\bar{D}/Q + \lambda\, Eg_Q(X)\bar{D}/Q,$$

where

a is the holding cost per unit time including allowance for interest, depreciation, obsolescence, and physical costs of maintaining inventory,

c is the variable cost per unit of the item in question,

Q is the order quantity,

A is the cost of placing an order including any set-up cost or other identifiable fixed cost of procurement,

$\overline{D}$ is average or expected demand,

E is the item essentiality, and

g_o is the expected number of unit-weeks of shortage per order cycle.

For any assigned λ, this objective function can be minimized for each item. Operating at this minimum for all items will generate an average level of shortages and average costs that can be extimated by simulation. The expected costs for operation of the whole system can be estimated by summation over all items or from an adequate sample of items to generate costs as a function of λ. The parameter λ may then be determined to fit any assigned budget. It may be further interpreted, however, as the reciprocal of the (maximum) reduction in shortages that could be provided by a dollar increase in budget.

The actual determination of X and Q as functions of λ and the item parameters is effected by formulas derived from simulation with a simplified environmental model which takes rather drastic liberties with the observed demand distribution and with formulas relating thereto. The adequacy of this experimentally-achieved optimization must be repeatedly tested by pilot tests for which the simplifying assumptions are abandoned.

The system developed has been applied to two classes of electronics items, on a pilot basis at the Electronics Supply Office. Before this full pilot test, however, extensive use of computer simulation was employed to compare expected performance of the proposed system with standard practice. For the preliminary pilot tests, a sample of items was treated by using historical lead time distributions and actual past demands (rather than Monte Carlo distributions) and evaluating costs, shortages, and other performance characteristics.

For the system now in pilot test, two uses of simulation are employed on-line. One involves use of alternative values for λ as intermediate steps in preparing cost-effectiveness curves. Choice of λ is based on these curves and on available budgets. For the measure of effectiveness another simulation is required since essentiality-weighted unit-weeks of shortage are not now tallied—the conventional not-in-stock (NIS) taillies do not consider duration of shortage, essentiality weights, or number of units short, but merely tally line items from requisitions that encounter not-in-stock conditions. Computer simulation is necessary in order to estimate the shortages being generated in the system. The measure of effectiveness is itself available only from on-line simulation.

A somewhat more complicated situation arises if separate budgets are involved for holding costs and for order costs so that the parameter A, the cost of placing an order (or some system parameter which raises all item order costs a fixed amount or in a fixed proportion), is also treated as adjustable. The parameter A could be adjusted to keep total system order costs within an assigned budget or to keep the average number of orders placed within the capacity provided by the staff of personnel available for procurement actions. Simultaneous adjustment of two parameters A and λ to match two independent constraints leads to a more complicated on-line simulation problem, but presents no theoretical difficulties.

AN AIRLIFT EXAMPLE

The second example for which on-line simulation could prove useful arises in connection with a somewhat generalized airlift staging problem. This was concerned with planning the movement of specified weights (or cubage) of cargo of various types from a number of specified origins to a number of specified destinations by using any of a set of specified links or routes. Transloading operations and fueling stops were admissible. Marshalling schedules for aircraft were also specified. The objective function was the time to complete the airlift which was to be minimized by using the least number of planes sufficient to achieve the minimum time. No stochastic elements were included.

This simulation was designed primarily for experimentation with alternative strategies for cargo prepositioning and aircraft marshalling, which provide inputs to the computer runs. The computer program could be used, however, in the course of an actual airlift to react to a real situation as it develops. Random variations in operating times, aborts, and other unforeseen circumstances would lead to divergence from the anticipated schedule, and a new set of runs could be made to revise schedules to minimize time required to complete the airlift.

Both examples, the stock control programs and the airlift programs, envisage use of simulation in experimentation, pilot test, and on-line control. As examples of simulation in on-line control they are extremely simple. The stock level case is concerned only with appropriate setting of one or more Lagrangian multipliers, say λ and possibly A, to adjust to current budgets. The airlift case is concerned only with examination of permissible modifications of the unexecuted marshalling schedules. These examples suffice to illustrate something of the general nature of on-line simulation, the needs it serves, and the manner in which it may be used.

The term on-line simulation suggests a requirement for computation in real-time, and this in turn suggests ultra high-speed computation. In the two examples cited, as in many more, real-time operation does not turn out to impose any serious speed requirement. In the safety level case, many of the Supply Demand Control Points currently operate on the basis of quarterly stock status reporting and quarterly review. In no case are decisions required oftener than on a weekly or semi-monthly basis. Simulation of the course of the supply system for the balance of the budget year or possibly one future year suffices for all practical purposes. In the airlift case it would probably suffice in practice to update the marshalling schedules (which represent the primary strategic decision variables open to modification) every eight hours after the airlift gets underway. In other examples one might cite from industry, say process control, or from military applications, say combat coordination center operation, the real-time requirement may be more drastic.

On-line simulation further suggests the existence of direct on-line data links or control channels automatically connecting the computer to the control settings of an operating system.

In the extremely simple cases we have described above, the very relaxed time requirements provide ample time for human intervention and even human decision making in response to computer results from on-line simulation. Even in the more continuous cases of process control, air traffic control, or combat coordination, similar use of human communication links and human decision making is the rule rather than the exception.

On-line simulation finally suggests an extensive amount of experimental simulation, perhaps with many Monte Carlo runs. In the examples cited, as well as in practice generally, the on-line simulations are required for selection among a relatively narrow range of alternatives with the simulation greatly aided by previous analytical results. The simulation program reduces, in most cases, to a relatively streamlined evaluator of the aggregate effects of different decisions selected from restricted classes of decision alternatives. The classic case is perhaps the chess playing programs which evaluate the relative strength of alternative chess situations one or two moves ahead, and optimize the current move with respect to this limited planning horizon. Most use of on-line simulation will of necessity be of modest scope for reasons of time, economy, and rapidly diminishing returns from increasing complexity.

The ever increasing size and speed of modern computers, the growing complexity of many control problems, and the gradual realization of the potential savings from timely and efficient control will bring about a greater use of on-line simulation for control purposes. Reliance on experience and intuition is yielding to greater dependence on experiment both with and without computer simulation. As numerical experimentation grows more common, the results of experiments in the design phase of a new system will become harder to interpret. In the face of normal uncertainty an optimum strategy for a long planning period may prove to be very suboptimal in comparison with a flexible strategy. The hazards of relying on either intuition or a fixed policy will justify major effort to refine or update decision making on-line. The best general procedure for this appears to be some form of on-line simulation specially designed for control of the final operating system.

* * *

GENERAL SIMULATION MODEL FOR LOGISTICS OPERATION IN A RANDOMLY DAMAGED SYSTEM*

John E. Walsh

System Development Corporation
Santa Monica, California

The logistics operation considered is imbedded in a system that is subject to damage. Natural attrition and induced attrition can both occur. This system is concerned with the availability and/or use of several different types of items during a specified period of time. The purpose of logistics is to furnish transportation, reduce natural attrition through maintenance, and return damaged items to operational condition. The capability of the logistics operation is examined by simulating the operation of the entire system for the specified time period. The damage received by the various parts of the system depends on the situation at the start of the time period, the use made of the items, the nature of the induced attrition, and the maintenance procedures, etc. Attrition occurs on a probability distribution basis and is introduced by a Monte Carlo procedure. The time required to return a damaged item to operational condition depends on the type of damage, the damage received by other items, the characteristics of the logistics operation, the damage received by the logistics part of the system, etc. To obtain the simulation model, the over-all system is divided into subsystems on the basis of the use and location of the items, the properties of the logistics operation, etc. The time scale used is discrete; i.e., all important effects are considered to occur during one of a specified finite set of time intervals. The same form of simulation model is used for representing the operation of each subsystem for each time interval. That is, the mathematical model has the same functional form, but the variables in this functional form can differ with the subsystem and time interval. These subsystems can interact in many ways, e.g., items can be transferred between subsystems, damage received in one subsystem can affect the logistics operations in other subsystems, etc. By using time as the basis for coordination among subsystems, the mathematical model and random assignments for attrition can be programmed on a high-speed computer. The simulation is performed by obtaining first the operational results for all subsystems at the first time considered; then, by using the first-time results, the second-time results are obtained, etc. Examination of the subsystem results for each time interval should furnish a good indication of the capabilities of the logistics operation. Repetitions of this simulation furnish an indication of the effects of statistical variation.

INTRODUCTION

In practice, logistics operations are virtually always parts of systems which are devoted to the accomplishment of one or more specified purposes. Then the suitability of a given logistics operation, or the relative desirabilities of several competitive logistics operations, ordinarily cannot be determined independently of the system characteristics. In fact, the worth of a logistics operation seems to be most meaningfully evaluated by consideration of the behavior of the system when this logistics operation is used.

---

*Prepared as System Development Corporation Paper SP-116.

When a system is extensive and complicated, a deterministic examination of its detailed operation is seldom feasible, even by the use of simulation and high-speed computers. If allowance is also made for random effects, such as damage, the difficulty of examining the system operation in detail is greatly increased. However, by suitable aggregation of detail, so that only the important effects are considered, the principal characteristics of the operation of a system can often be simulated in a feasible manner, even when random damage occurs. The purpose of this paper is to present a model of a somewhat general nature for performing simulations of this type.

For the simulation model presented, the system in which the logistics operation is imbedded is not of a completely general nature. The viewpoint adopted is that the purpose of the simulation is to examine the logistics operation under realistic circumstances. This purpose can often be accomplished to a reasonable extent by limiting consideration to systems which are concerned with the use and availability of various types of items and for which the "strategies" for system operation are restricted to transfer of items within the system and to maintenance and repair capabilities. Here natural and induced damage to the items are the only random effects that are introduced, and these depend on probability distributions which, during the time period considered, are influenced only by happenings within the system.

The major loss of generality in the model presented is that, during the time period considered, the probability distributions for induced attrition are influenced only by occurrences within the system containing the logistics operation. Often the induced attrition arises from a separate system. Then the interactions between the two systems during the time interval, and their effect on the operation strategies adopted by the systems, can be important in evaluating the behavior of the system containing the logistics operation. However, when the principal interest is in the logistics operation, the restricted system considered in this paper should be satisfactory in most cases. Actually, by judicious specification of the functional forms for the probability distributions of induced attrition, the effects of the second system can usually be at least roughly approximated. When a realistic evaluation of the behavior of the entire system is desired, a separate simulation model can be developed for the system influencing the induced attrition. Then the combination of the two systems can be simulated by relating the two models through their strategies and the probability distributions for induced damage.

In the model, damage can occur as the result of induced attrition and/or natural attrition. As opposed to induced attrition, which arises from unusual causes and can be deliberately introduced by another system, natural attrition represents the random damage that occurs in the ordinary day-to-day operation of the system.

A characteristic of the simulation model presented is that some of the strategies (those involving transfer of items) can be chosen on the basis of an estimation of the future behavior of the system. That is, suppose that the system operation has been simulated up to the end of a given time interval; by using these observed results, the future system behavior is anticipated by assuming that the attrition levels that have been observed are the actual attrition levels for future time intervals. This procedure generalizes the less realistic procedure of fixing these strategies prior to the simulation and includes the fixed strategy situation as a special case.

A fundamental part of the development of the simulation model is the conversion of the continuous situation for system operation to an approximately equivalent discrete situation which is feasible for simulation on a high-speed computer. This is accomplished by dividing

the time period during which simulation is to be performed into time intervals of equal length and by dividing the system into appropriate subsystems. The selection of the subsystems and the basic time interval depends on the system considered and is determined on a judgment basis.

Every subsystem is concerned with the operation and the transportation of items. Some of the subsystems are also concerned with the performance of maintenance and/or the repair of damage for items. All operations of the system occur as a result of the use of the items in the system. These operations include the logistics functions of transportation, maintenance, and repair; in fact, some of the items are introduced exclusively for the logistics part of the system. Also some of the subsystems are included for special purposes; that is, some subsystems are introduced for the performance of maintenance and/or repair, while others are included for the introduction or removal of items from the system.

Starting with the initial condition of the system (specified), the simulation is performed over all subsystems for the first time interval. Here, and at all steps, the damage actually received is introduced by a Monte Carlo procedure. On the basis of the initial conditions and the results at the end of the first time interval, the simulation is conducted for the second time interval. In a stepwise fashion, this procedure is continued until the condition of the system at the end of the final time interval is determined. The totality of results, for all the time intervals, constitutes a simulation of the behavior of the system. Several independent repetitions of this simulation should furnish a good indication of the statistical variation in system operation that is due to the random character of the damage received.

In the mathematical model, each item that occurs in the system operation is considered separately. Roughly stated, the breakdown used for the model is made with respect to the time interval considered, the item considered, the type of item, the subsystem location of the item, the status (e.g., in operation, in transit to repair, in maintenance) of the item, the maintenance schedule for the item, and the anticipated future-location scheduling for the item. Of these, the time interval considered and the location for the item considered receive the principal attention in the model. In fact, as indicated by the relationships that are used to state the mathematical model, the same form of simulation model can be used for each combination of time interval, item, and item location. That is, with little loss in generality, the mathematical model can be considered to have a fixed functional form which involves variables that depend on the time interval, item, and item location. Having fixed functional forms of this nature is convenient from the viewpoint of simulation on a high-speed computer. Namely, the problem is reduced to that of determining the values of the variables that occur in the fixed functional forms.

The specification of the mathematical model is made in general terms. That is, the available values that are sufficient to determine a new value are always explicitly identified; however, the function of these values that furnishes the new value is seldom explicitly stated. The model is sufficiently general to represent many situations of interest but does not concern itself with the problem of explicitly determining the various functions that occur in its statement. For each given system to be simulated, this problem would ordinarily be solved by a technical examination of the properties and the purposes of the system.

The model presented represents an appreciable aggregation of detail. However, in spite of this consolidation, a very large number of evaluations must be made for each time interval. In fact, unless care is exercised in selecting the basic time interval, the subsystems, the types of items, the functional forms, the levels of maintenance considered, the levels of damage

that can be received, etc., even the model presented may be too massive for simulation on any presently available high-speed computer.

Simulations of the type considered seem to be most valuable for developing an understanding of a logistics operation that is imbedded in a complicated system. These simulations should also be useful in examining the relative desirability of stated logistics strategies and the relative desirability of specified logistics operations.

The next section (Remarks Concerning the Model) contains statements of a more detailed nature about the characteristics of the simulation model. The quantities that occur in the model are identified and defined in the section in sequence, which is titled "Definition of Notation." The next to last section ("Assumptions and Conditions"), embodies statements concerning limitations of the model and about some of its properties. The final section ("Functional Relationships and Mathematical Model") presents the statement of the general simulation model.

REMARKS CONCERNING THE MODEL

The mathematical model presented for simulation uses involves so many different quantities and is of such a complicated nature that no attempt will be made to give a detailed word description of all its properties. The only description given that covers all the details of the model is of a technical character, and this is presented in the next three sections of the paper. However, discussions do seem to be warranted for some of the properties of the model and these discussions are given in this section.

In addition to the transfer schedules, various other quantities in the model have the characteristics of "strategies;" that is, these quantities have influence on the system operation and, at least to some extent, can be deliberately specified. Some of these strategies are included in the condition of the system at the beginning of the time period. Some of the others are represented by the explicit statements of the functions that are used in parts of the mathematical model; that is, these functional forms are determined by the characteristics of the system, and improvement of these characteristics is one method of improving the system operation. For example, the functional forms for the times spent in transit, in maintenance, and in repair represent strategies of this type. The maintenance schedules specified for the items represent strategies of another type, since they influence the probability distributions for natural attrition. Investigation of the effects of the strategies on the system behavior is perhaps the most important reason for simulating the system operation.

The maintenance schedule for an item is expressed in terms of the accumulated amount of operational time for that item (accumulated from the beginning of the time period being simulated). However, an interval of time during which an item is in operational condition does not necessarily represent a full interval of operational time for that item. In fact, the amount of operational time accumulated during a time interval in which an item is in operational condition can depend on the time interval considered, the subsystem location of the item, the item type, whether the item is in transit, the number of other items of this type in this subsystem at this time, etc. An item accumulates no operational time during a time interval if it is not in operational condition both at the beginning and at the end of this interval. Although accumulated operational time is not necessarily expressed as an integral number of time intervals, the maintenance schedules are expressed in this fashion. An item which was in an operational status at the beginning of a time interval needs maintenance at the end of this interval if its accumulated operational time equals or exceeds a critical value in the maintenance schedule at the end of the time interval.

At the end of each time interval, the transfer schedules that are to be introduced for items are chosen by anticipating the future behavior of the system. Since the system operation for future time intervals also involves the selection of transfer schedules at these future times, the amount of computation required in estimating the future behavior can easily reach an infeasible level. In many cases, adopting specified schedules for the items that are not used in the logistics operation may be desirable. Restricting the opportunity to make schedule changes to every third or fourth time interval represents another possibility for reducing the amount of computation. An important part of the mathematical model is an efficiency criterion for simultaneously deciding, at the end of each time interval, which combination of the eligible transfer schedules is to be used. A satisfactory specification of this criterion, for each time interval, represents a difficult problem that must be solved for each system that is simulated.

For items of a given type, the damage that can be received by an item is represented by a finite set of levels; also, a common set of levels is used for representing the kinds of maintenance that an item can be scheduled to undergo. To simplify the presentation, the number of damage levels is considered to be the same for all types of items and the number of maintenance levels is taken to be the same for all types of items. This simplification results in no loss of generality since not all of the levels need to be used for every type of item. In general, the meanings of the levels may differ greatly from one type of item to another. The procedure of using levels is evidently suitable for maintenance and should nearly always furnish an acceptable representation of damage, if enough damage levels are used.

At the end of any given time interval, the level of future natural damage and the level of future induced damage each can be estimated as the average of the observed values over all the time intervals up to and including the given interval. Here the estimates are for the average damage during a single time interval and are obtained for every combination of item type and subsystem. Then, the anticipated future behavior of the system is predicted by starting with the situation at the end of the given time interval and then simulating the future system operation under the assumption that the observed average values equal the actual damage values for each of the future time intervals. Here probability enters only in a very restricted sense, and is introduced by a Monte Carlo process. Specifically, let an average attrition value be stated in the form $I + f$, where I is a nonnegative integer and f is a nonnegative proper fraction; then the damage level actually used is selected on a Monte Carlo basis and has the value I with probability $1 - f$ and the value $I + 1$ with probability f. This random rounding of an average damage value to one of the damage levels considered can be done (independently) for each combination of item and future time interval. In this manner, a definite damage level is determined for each item at each future time interval.

For a given subsystem, a different type of logistics item might be needed for each combination of logistics task and item type. Sometimes, however, one type of logistics item performs the same logistics function for items of several types. Consequently, groupings of items of several different types can occur in the determination of the effects of numbers of items on the logistics operation for a subsystem.

Finally, the wide range of application for the simulation model presented should probably be emphasized. The principal restrictions imposed by the mathematical model are of a general nature; namely, statements of which available values are sufficient for the determination

of a new value. Thus the model is applicable for the simulation of any system for which a basic time interval, a set of subsystems, etc., can be determined so that the relations actually satisfied (to a reasonable approximation) depend on the same sets of variables as those for the corresponding relations in the model. Of course, if only a subset of the set of available values specified by the model is needed for determining the new value, the relation stated in the mathematical model is satisfied. Although cost considerations are not explicitly introduced, they can appear in the initial conditions, the transfer schedules, the efficiency criterion, the functional relationships, etc.

DEFINITION OF NOTATION

Extensive use is made of symbolism in presenting the functional form of the mathematical model for a specified subsystem at a given time. The definitions for this notation are given below. These definitions should also be useful in identifying the factors that are explicitly considered in the simulation model presented.

t = time interval considered ($t = 1, \ldots, T$).

u = identification number for an item ($u = 1, \ldots, U$).

v_u = classification type for u-th item ($v_u = 1, \ldots, V$). Here types $1, \ldots, L$ refer to nonlogistics items while types $L + 1, \ldots, V$ refer to logistics items.

$j_u(t)$ = subsystem location of u-th item during the t-th time interval. The subsystems are identified by the numbers $1, \ldots, J$; $j_u(0)$ is specified initial condition.

$s_u(t)$ = status of u-th item at end of t-th time interval; $s_u(0)$ is specified initial condition. Seven different status levels are considered:
Level 1 ≡ item in operation (not in transit);
Level 2 ≡ item operational but in transit;
Level 3 ≡ item in transit to receive maintenance;
Level 4 ≡ item in transit for repair of damage that is at different level from that present at end of previous time interval;
Level 5 ≡ item in transit for repair of damage that is at same level as that present at end of previous time interval;
Level 6 ≡ receiving maintenance;
Level 7 ≡ receiving repair.

$T_O^{(u)}(t)$ = accumulated amount of operational time at end of t-th interval; $T_O^{(u)}(0) = 0$.

$m_u^{(s)}$ = maintenance schedule for u-th item; depends on accumulated amount of operational time for that item.

$M_u[T_O^{(u)}(t)]$ = maintenance level needed by u-th item at end of t-th time interval. Maintenance levels considered are $0, 1, \ldots, M$ where level 0 denotes that no maintenance needed.

$d_N^{(u)}(t)$ = level of damage from natural attrition received by u-th item during t-th time interval (in addition to induced damage received in this time interval and damage present at beginning of interval). Levels considered for damage from any source are $0, 1, \ldots, D$, where level 0 denotes that no damage received.

$d_I^{(u)}(t)$ = level of damage from induced attrition received by u-th item during t-th time interval (in addition to natural damage received in this time interval and damage present at beginning of interval).

$P_N^{(u)}(d;t)$ = probability that u-th item will receive natural damage of level d during t-th time interval (d = 0, 1, . . . , D). The value of $d_N^{(u)}(t)$ is a random choice from this discrete probability distribution.

$P_I^{(u)}(d;t)$ = probability that u-th item will receive induced damage of level d during t-th time interval (d = 0, 1, . . . , D). The value of $d_I^{(u)}(t)$ is a random choice from this discrete probability distribution.

$D_N^{(u)}(t, j)$ = arithmetical average of $d_N^{(u')}(t')$ over all $t' \leq t$ and all u' such that $v_{u'} = v_u$ and $j_{u'}(t') = j$.

$D_I^{(u)}(t, j)$ = arithmetical average of $d_I^{(u')}(t')$ over all $t' \leq t$ and all u' such that $v_{u'} = v_u$ and $j_{u'}(t') = j$.

$d_u(t)$ = over-all damage level of u-th item at end of t-th time interval; $d_u(0)$ is specified initial condition.

$t_o^{(u)}(t)$ = anticipated number of additional time intervals, after t-th, required for u-th item to become operational; here $t_o^{(u)}(t) = 0$ implies that $s_u(t) = 1$ or 2. Determined by starting with situation at end of t-th interval and simulating results for future time intervals under assumption that $d_N^{(u')}(t') = D_N^{(u')}[t, j_{u'}(t')]$ and $d_I^{(u')}(t') = D_I^{(u')}[t, j_{u'}(t')]$ for $t' > t$ and all u'; set equal to T + 1 - t if item not anticipated to be in operation by end of T-th interval.

$t_C^{(u)}(t)$ = anticipated number of additional time intervals, after t-th, required for u-th item to complete transit (change of subsystems); here $t_C^{(u)}(t) = 0$ implies that u-th item not in transit at end of t-th interval. Determined by starting with situation at end of t-th interval and simulating results for future time intervals, under the assumption that $d_N^{(u')}(t') = D_N^{(u')}[t, j_{u'}(t')]$ and $d_I^{(u')}(t') = D_I^{(u')}[t, j_{u'}(t')]$ for $t' > t$ and all u'; set equal to T + 1 - t if transit is not anticipated to be completed by end of T-th interval.

$t_m^{(u)}(t)$ = anticipated number of additional time intervals, after t-th, required for maintenance to be completed for u-th item; here $t_m^{(u)}(t) = 0$ implies that u-th item is not receiving maintenance at end of t-th interval. Determined by starting with situation at end of t-th interval and simulating results for future time intervals under the assumption that $d_N^{(u')}(t') = D_N^{(u')}[t, j_{u'}(t')]$ and $d_I^{(u')}(t') = D_I^{(u')}[t, j_{u'}(t')]$ for $t' > t$ and all u'; set equal to T + 1 - t if maintenance is not anticipated to be completed by end of T-th interval.

$t_r^{(u)}(t)$ = anticipated number of additional time intervals, after t-th, required for repair to be completed for u-th item; here $t_r^{(u)}(t) = 0$ implies

that u-th item is not receiving repair at end of t-th interval. Determined by starting with situation at end of t-th interval and simulating results for future time intervals under the assumption that $d_N^{(u')}(t') = D_N^{(u')}[t, j_{u'}(t')]$ and $d_I^{(u')}(t') = D_I^{(u')}[t, j_{u'}(t')]$ for $t > t'$ and all u'; set equal to $T + 1 - t$ if repair is not anticipated to be completed by end of T-th interval.

$j_m^{(u)}(t)$ = subsystem in which u-th item is receiving maintenance during t-th time interval; here $j_m^{(u)}(t) = 0$ implies that u-th item is not receiving maintenance during t-th interval.

$j_r^{(u)}(t)$ = subsystem in which u-th item is receiving repair during t-th time interval; here $j_r^{(u)}(t) = 0$ implies that u-th item is not receiving repair during t-th interval.

$n_1(t, v, j)$ = number of operational items of type v (status 1 or 2) in subsystem j at end of t-th time interval which were in status 1 at beginning of this interval.

$B_v(t, j)$ = lower bound for number of items of type v that are desired to be in operation (status 1) in subsystem j at end of t-th time interval.

$n_s(t, v, j_1, j_2)$ = number of items of type v and in operation (status 1) in subsystem j_1 at end of t-th time interval that are scheduled to begin transit to subsystem j_2 at end of t-th interval; schedule can be stated in terms of $n_1(t, v, j)$. Several possible schedules can be specified for $n_s(t, v, j_1, j_2)$ with the schedule selected being determined by the anticipated future behavior of the system for each possibility, under the assumption that $d_N^{(u)}(t') = D_N^{(u)}[t, j_u(t')]$ and $d_I^{(u)}(t') = D_I^{(u)}[t, j_u(t')]$ for $t' > t$ and all u. These possible schedules are denoted by $n_s^{(k_S)}(t, v, j_1, j_2)$, $k_S = 1, \ldots,$ $K_S(t, v, j_1, j_2)$.

$n_a(t, v, j_1, j_2)$ = number of items of type v and in operation (status 1) in subsystem j_1 at end of t-th time interval that actually begin transit to subsystem j_2 at end of t-th interval.

$q_u[t, v_u, j_u(t), j_2]$ = queuing priority assigned u-th item in selection of which items of type v_u and status 1 at end of t-th time interval are to start transit from subsystem $j_u(t)$ to subsystem j_2 at end of t-th interval; $q_u[t, v_u, j_u(t), j_2]$ is positive for $s_u(t) = 1$ and zero otherwise.

$S_O(t, v, j)$ = specification of the in-transit subsystem locations, in given order, for an operational item (status 2) of type v that began transit from subsystem j and at end of t-th time interval; last of these subsystems is destination of item. The item changes from status 2 to status 1 when the destination subsystem, denoted by $J_O(t, v, j)$, reached in operational condition. Several alternative specifications of in-transit subsystem locations could be given, the specification selected being determined by the anticipated future behavior of the system under the assumption that $d_N^{(u)}(t') = D_N^{(u)}[t, j_u(t')]$ and $d_I^{(u)}(t') = D_I^{(u)}[t, j_u(t')]$ for $t > t'$ and all u. These alternatives are denoted by $S_O^{(k_O)}(t, v, j)$, $k_O = 1, \ldots, K_O(t, v, j)$.

$S_M(t, v, m, j)$ = specification of the in-transit subsystem locations, in given order, for an item of type v needing maintenance at level m which began transit from subsystem j, and at end of t-th time interval; last of these subsystems is destination of item for receiving maintenance. The item changes from

status 3 to status 6 when the destination subsystem is reached in undamaged condition. Several alternative specifications of in-transit subsystem locations could be given with specification selected being determined by the anticipated future behavior of the system, under the assumption that $d_N^{(u)}(t') = D_N^{(u)}[t, j_u(t')]$ and $d_I^{(u)}(t') = D_I^{(u)}[t, j_u(t')]$ for $t > t'$ and all u. These alternatives are denoted by $S_M^{(k_M)}(t, v, m, j)$, $k_M = 1, \ldots, K_M(t, v, m, j)$.

$S_R(t, v, d, j)$ = specification of the in-transit subsystem locations, in given order, for an item of type v needing repair of level-d damage which began transit for repair of this level of damage from subsystem j and at end of t-th time interval. Several alternative specifications of in-transit subsystem locations could be given with specification selected being determined by the future behavior of the system under the assumption that $d_N^{(u)}(t') = D_N^{(u)}[t, j_u(t')]$ and $d_I^{(u)}(t') = D_I^{(u)}[t, j_u(t')]$ for $t > t'$ and all u. These alternatives are denoted by $S_R^{(k_R)}(t, v, d, j)$, $k_R = 1, \ldots, K_R(t, v, d, j)$.

$e(t)$ = efficiency criterion for simultaneously selecting the strategies $n_S(t, v, j_1, j_2)$, $S_O(t, v, j)$, $S_M(t, v, m, j)$, and $S_R(t, v, d, j)$ from the sets of possibilities $\{n_S^{(k_S)}(t, v, j_1, j_2)\}$, $\{S_O^{(k_O)}(t, v, j)\}$, $\{S_M^{(k_M)}(t, v, m, j)\}$, and $\{S_R^{(k_R)}(t, v, d, j)\}$. Determined by starting with situation at end of t-th time interval and simulating results for future time intervals under the assumption that $d_N^{(u)}(t') = D_N^{(u)}[t, j_u(t')]$ and $d_I^{(u)}(t') = D_I^{(u)}[t, j_u(t')]$ for $t' > t$ and all u. Here, for $t < t' \le t + t_e(t)$, the $n_S(t', v, j_1, j_2)$, $S_O(t', v, j)$, $S_M(t', v, m, j)$, and $S_R(t', v, d, j)$ are determined from $e(t')$; in general, all possible selections for these strategies must be considered. For $t' > t + t_e(t)$, however, each of these strategies is uniquely specified on the basis of an efficiency criterion $e[t', t, t_e(t)]$ that is determined by the actual and anticipated results (all possible combinations of selections) for the situations existing at the end of time intervals up to and including $t + t_e(t)$.

$g_C(v, d, m, j)$ = type of logistics item that is used for transportation in subsystem j of in-transit items of type v, damage level d, and needing maintenance of level m. Here in-transit items are grouped with respect to type, damage level, maintenance level needed, and subsystem location in such a manner that all items of a group can be handled by the same type of logistics item; hence many combinations of values for v, d, m, j may yield the same value for $g_C(v, d, m, j)$. Note that one or both of d and m are zero.

$g_M(v, m, j)$ = type of logistics item used for maintenance of items of type v that are receiving maintenance at level m in subsystem j.

$g_R(v, d, j)$ = type of logistics item used for repair of items of type v that are receiving repair of level d damage in subsystem j.

$N_C(t, t', g_C, j)$ = anticipated number of in-transit items in subsystem j at the end of each of the time intervals $t + 1, \ldots, t + t' \le T$ that are transported in this subsystem by logistics items of type g_C. This vector is determined by starting with the situation at the end of t-th interval and simulating results for time intervals $t + 1, \ldots, t + t'$ under the assumption that $d_N^{(u)}(t'') =$

$D_N^{(u)}[t, j_u(t'')]$ and $d_I^{(u)}(t'') = D_I^{(u)}[t, j_u(t'')]$ for $t + 1 \le t'' \le t + t'$ and all u.

$N_M(t, t', g_M, j)$ = anticipated number of items in maintenance in subsystem j at the end of each of the time intervals $t + 1, \ldots, t + t' \le T$ that are handled in maintenance in this subsystem by logistics items of type g_M. This vector is determined by starting with the situation at the end of t-th interval and simulating results for time intervals $t + 1, \ldots, t + t'$ under the assumption that $d_N^{(u)}(t'') = D_N^{(u)}[t, j_u(t'')]$ and $d_I^{(u)}(t'') = D_I^{(u)}[t, j_u(t'')]$ for $t + 1 \le t'' \le t + t'$ and all u.

$N_R(t, t', g_R, j)$ = anticipated number of items in repair in subsystem j at the end of each of the time intervals $t + 1, \ldots, t + t' \le T$ that are handled in repair in this subsystem by items of type g_R. This vector is determined by starting with the situation at the end of t-th interval and simulating results for time intervals $t + 1, \ldots, t + t'$ under the assumption that $d_N^{(u)}(t'') = D_N^{(u)}[t, j_u(t'')]$ and $d_I^{(u)}(t'') = D_I^{(u)}[t, j_u(t'')]$ for $t + 1 \le t'' \le t + t'$ and all u.

$N(t, t', g, j)$ = anticipated number of logistics items of type g that are in operation (status 1) in subsystem j at the end of each of the time intervals $t + 1, \ldots, t + t' \le T$. This vector is determined by starting with the situation at the end of t-th interval and simulating results for time intervals $t + 1, \ldots, t + t'$ under the assumption that $d_N^{(u)}(t'') = D_N^{(u)}[t, j_u(t'')]$ and $d_I^{(u)}(t'') = D_I^{(u)}[t, j_u(t'')]$ for $t + 1 \le t'' \le t + t'$ and all u.

$t_O[t, v, t^{(o)}, j^{(o)}, j_O(t)]$ = anticipated additional number of time intervals required, after t-th interval, for operational item of type v that began transit at the end of time interval $t^{(o)}$ from subsystem $j^{(o)}$ and which is now in subsystem $j_O(t)$ to complete transit in $j_O(t)$. This is determined on basis of $N_C(t, t', g_C, j)$ and $N(t, t', g, j)$; value of $t_O[0, v, t^{(o)}, j^{(o)}, j_O(0)]$ is the specified initial condition. Set equal to $T + 1 - t$ if transit in $j_O(t)$ is not anticipated to be completed by end of T-th interval. Item leaves a transit subsystem and enters next subsystem specified by $S_O(t^{(o)}, v, j^{(o)})$ when $t_O[t, v, t^{(o)}, j^{(o)}, j_O(t)] = 0$.

$j_O(t, v, t^{(o)}, j^{(o)})$ = subsystem location during t-th time interval for operational unit (status 2) of type v that began transit at the end of time interval $t^{(o)}$ from subsystem $j^{(o)}$.

$t_M[t, v, m, t^{(M)}, j^{(M)}, j_M(t)]$ = anticipated additional number of time intervals required, after t-th interval, for item of type v that needs maintenance of level m, which began transit at end of time interval $t^{(M)}$ from subsystem $j^{(M)}$, and is now in subsystem $j_M(t)$, to complete transit in $j_M(t)$. This is determined on the basis of $N_M(t, t', g_M, j)$ and $N(t, t', g, j)$; value of $t_M[0, v, m, t^{(M)}, j^{(M)}, j_M(0)]$ is the specified initial condition. Set equal to $T + 1 - t$ if transit in $j_M(t)$ is not anticipated to be completed by end of T-th interval. Item leaves a transit subsystem and enters

next subsystem specified by $S_M(t^{(M)}, v, m, j^{(M)})$ when $t_M[t, v, m, t^{(M)}, j^{(M)}, j_M(t)] = 0$.

$j_M(t, v, m, t^{(M)}, j^{(M)})$ = subsystem location during t-th time interval for item of type v that needs maintenance of level m and began transit at end of time interval $t^{(M)}$ from subsystem $j^{(M)}$.

$t_R[t, v, d, t^{(R)}, j^{(R)}, j_R(t)]$ = anticipated additional number of time intervals required, after t-th interval, for item of type v that needs repair of damage level d, which began transit for repair of this level of damage at the end of time interval $t^{(R)}$ from subsystem $j^{(R)}$, and is now in subsystem $j_M(t)$, to complete transit of $j_M(t)$. Determined on the basis of $N_R(t, t', g_R, j)$ and $N(t, t', g, j)$; the value of $t_R[0, v, d, t^{(R)}, j^{(R)}, j_R(0)]$ is the specified initial condition. Set equal to $T + 1 - t$ if transit in $j_R(t)$ is not anticipated to be completed by end of T-th interval. An item leaves a transit subsystem and enters the next subsystem specified by $S_R(t^{(R)}, v, d, j^{(R)})$ when $t_R[t, v, d, t^{(R)}, j^{(R)}, j_R(t)] = 0$.

$j_R(t, v, d, t^{(R)}, j^{(R)})$ = subsystem location during t-th time interval for an item of type v that needs repair of level d damage and which began transit for repair of damage at this level at the end of time interval $t^{(R)}$ from subsystem $j^{(R)}$.

$j_u^{(o)}(t)$ = location of u-th item during t-th time interval when this item is operational and in transit (status 2) at end of (t - 1)-th interval.

$t_u^{(o)}(t)$ = anticipated number of additional time intervals, after t-th, for completion of transit in subsystem $j_u^{(o)}(t)$ for u-th item, when this item is in status 2 at end of (t - 1)-th interval. Determined on basis of $t_o[t, v, t^{(o)}, j^{(o)}, j_o(t)]$.

$j_u^{(M)}(t)$ = location of u-th item during t-th time interval when this item is in transit to maintenance (status 3) at end of (t - 1)-th interval.

$t_u^{(M)}(t)$ = anticipated number of additional time intervals, after t-th, for completion of transit in subsystem $j_u^{(M)}(t)$ for u-th item, when this item is in status 3 at end of (t - 1)-th interval. Determined on basis of $t_M[t, v, m, t^{(M)}, j^{(M)}, j_M(t)]$.

$j_u^{(R)}(t)$ = location of u-th item during t-th time interval when this item is in transit to repair (status 4 or 5) at end of (t - 1)-th interval.

$t_u^{(R)}(t)$ = anticipated number of additional time intervals, after t-th, for completion of transit in subsystem $j_u^{(R)}(t)$ for u-th item, when this item is in status 4 or 5 at end of (t - 1)-th interval. Determined on basis of $t_R[t, v, d, t^{(R)}, j^{(R)}, j_R(t)]$.

$T_M(t, v, m, j)$ = anticipated number of time intervals to completion of maintenance for item of type v that begins maintenance of level m in subsystem j at the end of the t-th time interval. Determined on the basis of $N_M(t, t', g_M, j)$ and $N(t, t', g, j)$ for t' sufficiently large. Set equal to $T + 1 - t$ if maintenance is not anticipated to be completed by end of T-th interval.

$T_R(t, v, d, j)$ = anticipated number of time intervals to completion of repair for item of type v that begins repair of level d damage in subsystem j at the end of

the t-th time interval. Determined on the basis of $N_R(t, t', g_R, j)$ and $N(t, t', g, j)$ for t' sufficiently large. Set equal to T + 1 - t if repair is not anticipated to be completed by end of T-th interval.

$C_u^{(o)}(t)$ = function used for stating whether the u-th item is chosen to begin transit at the end of the t-th interval when it was in operation (status 1) at beginning of this time interval. Has value 0 if $s_u(t - 1) = 1$ but u-th item not chosen to begin transit; has value 1 if $s_u(t - 1) = 1$ and item chosen to begin transit; has value 2 otherwise.

$\tau_u^{(o)}(t)$ = most recent time interval, up to and including t-th, such that $C_u^{(o)}[\tau_u^{(o)}(t)] = 1$.

$\tau_u^{(M)}(t)$ = most recent time interval, up to and including t-th, such that $M_u\{T_o^{(u)}[\tau_u^{(M)}(t)]\} > 0$ but $M_u\{T_o^{(u)}[\tau_u^{(M)}(t) - 1]\} = 0$.

$\tau_u^{(R)}(t)$ = most recent time interval, up to and including t-th, such that $d_u[\tau_u^{(R)}(t)] - d_u[\tau_u^{(R)}(t) - 1] > 0$.

$\tau_m^{(u)}(t)$ = most recent time interval, up to and including t-th, such that $t_C^{(u)}[\tau_m^{(u)}(t) - 1] > 0$ or $M_u\{T_o^{(u)}[\tau_m^{(u)}(t) - 1]\} = 0$.

$\tau_r^{(u)}(t)$ = most recent time interval, up to and including t-th, such that $t_C^{(u)}[\tau_r^{(u)}(t) - 1] > 0$ or $d_u[\tau_r^{(u)}(t) - 1] = 0$.

$n_C(t, v, d, m, j)$ = number of items of type v, damage level d, and needing maintenance at level m, that are in transit in subsystem j at end of t-th time interval; $n_C(0, t, v, d, m, j)$ is the specified initial condition.

$n_M(t, v, m, j)$ = number of items of type v that are receiving maintenance of level m in subsystem j at end of t-th time interval; $n_M(0, v, m, j)$ is the specified initial condition.

$n_R(t, v, d, j)$ = number of items of type v that are receiving repair of damage at level d in subsystem j at the end of t-th time interval; $n_R(0, v, d, j)$ is the specified initial condition.

$n_o(t, v, j)$ = number of items of type v in operation (status 1) in subsystem j at end of t-th time interval; $n_o(0, v, j)$ is the specified initial condition.

ASSUMPTIONS AND CONDITIONS

The model developed must satisfy some assumptions if the simulation is to furnish an acceptable approximation to the operation of the system being considered. These assumptions, which are concerned with the approximation of a continuous situation by a discrete situation, can be stated as follows:

(a) The basic time interval chosen for use is small enough to yield a reasonable approximation to the situation of continuous time.

(b) The basic time interval and the subsystems are such that the random damage received in one subsystem can be considered independent of the random damage received in any other subsystem during the same time interval.

(c) The subsystems are uncomplicated enough for their required properties to be satisified to a reasonable approximation. That is, the use of these subsystems, in conjunction with the conditions and relationships specified by the model, furnishes an acceptable representation of the system behavior when the other assumptions hold.

Whether assumptions (a) through (c) are acceptable for a given-size basic time interval and a specified set of subsystems depends on the system being approximated and on the conditions and relationships imposed by the model.

The functional relationships used in specifying the mathematical model are stated in the next section. Separately, some conditions are adopted for the operation of the model. These conditions serve two purposes. First, some fundamental logistics priorities are specified. Second, the simulation model is simplified without much curtailment of its generality and applicability. The conditions imposed are:

1. After completion of maintenance or repair, an item is considered to be operational (status 1) for at least one time interval.
2. Repair of damage takes precedence over maintenance. When an item needing maintenance or receiving maintenance is damaged, the item first receives repair. The required maintenance is then performed as if it were first coming due in the subsystem where the repair was completed. Functionally, this is accomplished by slightly reducing the value of $T_0^{(u)}(t)$, which has a value which specifies that maintenance is due, when the item becomes damaged; then maintenance will immediately become due as the item becomes operational at the end of repair. Thus, the model is such that an item cannot simultaneously be damaged and need maintenance.
3. After a change of status because of the need for maintenance or the need for repair, an item is considered to be in transit (perhaps in the same subsystem) for at least one time interval before maintenance or repair is started.
4. For each combination of values for time interval, subsystem location, type of item, level of damage, and maintenance level needed, the scheduled transfer is to a single destination subsystem and by use of a single route (specified by the in-transit subsystems and their order in the transit).
5. If an item becomes due for maintenance or receives damage while in transit, it is treated as if it were an item (not in transit) of the subsystem where this status change occurred.
6. If a damaged item becomes further damaged, it is treated as if the totality of damage were just received in the subsystem where it is now located. Any repair already received for the previous damage level is not considered. However, the repair received is taken into consideration in deciding how much the values of $d_N^{(u)}(t)$ and $d_I^{(u)}(t)$ are to increase the value of $d_u(t - 1)$.
7. Changes in the characteristics of an item that occur during a given time interval do not have any effect on the other items until at least one time interval later.
8. An item stays in each of its subsystem locations for at least one time interval.

These conditions are used, sometimes only implicitly, in the statement of the functional relationships among the quantities occurring in the mathematical model.

FUNCTIONAL RELATIONS AND MATHEMATICAL MODEL

The mathematical model for the simulation consists of the statement of a method whereby knowledge of all the pertinent quantities up to and including the end of the $(t - 1)$-th time interval can be used to evaluate these quantities at the end of the t-th interval. This model, combined

with the given initial conditions for the system, can be used for simulating the system operation over all the time intervals considered. Here the pertinent quantities for each time interval are those defined under the Definition of Notation.

As already mentioned, the general simulation model presented is not of a detailed nature, with the other values that are used to determine a given quantity having only been identified in many cases. That is, in the statement of the model, the evaluation of a given quantity is often expressed in the form of an unspecified function of specified quantities at stated times. Thus the other values which are sufficient for evaluation of this quantity are stated, but the form of the function used in the evaluation is not always stated. Of course, for a simulation of any given system, these functions would be specified on the basis of the properties of the system considered.

This section fulfills a dual purpose. First, for each quantity considered, the other values that are sufficient for determination of this quantity are specified. Second, the order in which functional relations for quantities are presented furnishes a statement of the mathematical model for the simulation. That is, this order is such that the specified values used for determining a quantity are always previously evaluated on the basis of quantities which have already occurred in the ordering and/or quantities for preceding time intervals. In situations involving anticipated future behavior of the system, the simulation is projected into the future but is based exclusively on values that have already been determined.

The quantities $n_O(t, v, j)$, $n_C(t, v, d, m, j)$, $n_M(t, v, m, j)$, and $n_R(t, v, d, j)$ are determined by direct examination of the situation at the end of the t-th time interval, while $C_u^{(o)}(t)$ is determined by this examination and the value of $s_u(t - 1)$. The v_u, $m_u^{(s)}$, $g_C(v, d, m, j)$, $g_M(v, m, j)$, $g_R(v, d, j)$, and $M_u[T']$ do not depend on the time interval considered while the $B_v(t, j)$ are specified independently of the simulation. The values of T, U, V, L, J, M, and D are specified. Some of the other quantities were defined in a functional form under the Definition of Notation. The remaining quantities that occur in the simulation model depend on the following forms of functional relationships:

$$T_M(t, v, m, j) = T_M\{t, v, m, j; T_M(t-1, v, m, j), N_M[t-1, T_M(t-1, v, m, j), g_M(v, m, j), j], N[t-1, T_M(t-1, v, m, j), g_M(v, m, j), j]\}$$

$$T_R(t, v, d, j) = T_R\{t, v, d, j; T_R(t-1, v, d, j), N_R[t-1, T_R(t-1, v, d, j), g_R(v, d, j), j], N[t-1, T_R(t-1, v, d, j), g_R(v, d, j), j]\}$$

$$j_O(t, v, t^{(o)}, j^{(o)}) = j_O\{t, j_O(t-1, v, t^{(o)}, j^{(o)}), t_O[t-1, v, t^{(o)}, j^{(o)}, j_O(t-1, v, t^{(o)}, j^{(o)})], S_O(t^{(o)}, v, j^{(o)})\} \text{ for } t > t^{(o)} \text{ and equals } j^{(o)} \text{ for } t = t^{(o)}$$

$$t_O[t, v, t^{(o)}, j^{(o)}, j_O(t, v, t^{(o)}, j^{(o)})] = t_O\{t, v, t^{(o)}, j^{(o)}, j_O(t, v, t^{(o)}, j^{(o)}); t_O[t-1, v, t^{(o)}, j^{(o)}, j_O(t-1, v, t^{(o)}, j^{(o)})], N_C(t-1, t_O[t-1, v, t^{(o)}, j^{(o)}, j_O(t-1, v, t^{(o)}, j^{(o)})], g_C(v, 0, 0, j^{(o)}), j^{(o)}), N(t-1, t_O[t-1, v, t^{(o)}, j^{(o)}, j_O(t-1, v, t^{(o)}, j^{(o)})], g_C(v, 0, 0, j^{(o)}), j^{(o)})\}$$

$$j_M(t, v, m, t^{(M)}, j^{(M)}) = j_M\{t, j_M(t\text{-}1, v, m, t^{(M)}, j^{(M)}), t_M[t\text{-}1, v, m, t^{(M)}, j^{(M)}, j_M(t\text{-}1, v, m, t^{(M)}, j^{(M)})], S_M(t^{(M)}, v, m, j^{(M)})\} \text{ for } t > t^{(M)} \text{ and equals } j^{(M)} \text{ for } t = t^{(M)}$$

$$\begin{aligned} t_M[t, v, m, t^{(M)}, j^{(M)}, j_M(t, v, m, t^{(M)}, j^{(M)})] = t_M\{&t, v, m, t^{(M)}, j^{(M)}, j_M(t, v, m, t^{(M)}, j^{(M)});\\ &t_M[t\text{-}1, v, m, t^{(M)}, j^{(M)}, j_M(t\text{-}1, v, m, t^{(M)}, j^{(M)})],\\ &N_C(t\text{-}1, t_M[t\text{-}1, v, m, t^{(M)}, j^{(M)}, j_M(t\text{-}1, v, m, t^{(M)}, j^{(M)})],\\ &g_C(v, 0, m, j^{(M)}), j^{(M)}), N(t\text{-}1,\\ &t_M[t\text{-}1, v, m, t^{(M)}, j^{(M)}, j_M(t\text{-}1, v, m, t^{(M)}, j^{(M)})],\\ &g_C(v, 0, m, j^{(M)}), j^{(M)})\} \end{aligned}$$

$$j_R(t, v, d, t^{(R)}, j^{(R)}) = j_R\{t, j_R(t\text{-}1, v, d, t^{(R)}, j^{(R)}), t_R[t\text{-}1, v, d, t^{(R)}, j^{(R)}, j_R(t\text{-}1, v, d, t^{(R)}, j^{(R)})], S_R(t^{(R)}, v, d, j^{(R)})\} \text{ for } t > t^{(R)} \text{ and equals } j^{(R)} \text{ for } t = t^{(R)}$$

$$\begin{aligned} t_R[t, v, d, t^{(R)}, j^{(R)}, j_R(t, v, d, t^{(R)}, j^{(R)})] = t_R\{&t, v, d, t^{(R)}, j^{(R)}, j_R(t, v, d, t^{(R)}, j^{(R)});\\ &t_R[t\text{-}1, v, d, t^{(R)}, j^{(R)}, j_R(t\text{-}1, v, d, t^{(R)}, j^{(R)})], N_C(t\text{-}1,\\ &t_R[t\text{-}1, v, d, t^{(R)}, j^{(R)}, j_R(t\text{-}1, v, d, t^{(R)}, j^{(R)})],\\ &g_C(v, d, 0, j^{(R)}), j^{(R)}), N(t\text{-}1,\\ &t_R[t\text{-}1, v, d, t^{(R)}, j^{(R)}, j_R(t\text{-}1, v, d, t^{(R)}, j^{(R)})],\\ &g_C(v, d, 0, j^{(R)}), j^{(R)})\} \end{aligned}$$

$$j_u^{(o)}(t) = j_o\{t, v_u, \tau_u^{(o)}(t), j_u[\tau_u^{(o)}(t)]\}$$

$$t_u^{(o)}(t) = t_o\{t, v_u, \tau_u^{(o)}(t), j_u[\tau_u^{(o)}(t)], j_u^{(o)}(t)\}$$

$$j_u^{(M)}(t) = j_M[t, v_u, M_u\{T_o^{(u)}[\tau_u^{(M)}(t)]\}, \tau_u^{(M)}(t), j_u\{\tau_u^{(M)}(t)\}]$$

$$t_u^{(M)}(t) = t_M[t, v_u, M_u\{T_o^{(u)}[\tau_u^{(M)}(t)]\}, \tau_u^{(M)}(t), j_u\{\tau_u^{(M)}(t)\}, j_u^{(M)}(t)]$$

$$j_u^{(R)}(t) = j_R\{t, v_u, d_u[\tau_u^{(R)}(t)], \tau_u^{(R)}(t), j_u[\tau_u^{(R)}(t)]\}$$

$$t_u^{(R)}(t) = t_R\{t, v_u, d_u[\tau_u^{(R)}(t)], \tau_u^{(R)}(t), j_u[\tau_u^{(R)}(t)], j_u^{(R)}(t)\}$$

$$j_u(t) = j_u(t-1) \text{ when } s_u(t-1) = 1, 6, \text{ or } 7$$
$$= j_u^{(o)}(t) \text{ when } s_u(t-1) = 2$$
$$= j_u^{(M)}(t) \text{ when } s_u(t-1) = 3$$
$$= j_u^{(R)}(t) \text{ when } s_u(t-1) = 4 \text{ or } 5$$

$$P_N^{(u)}(d;t) = P_N\{d;t, v_u, j_u(t), m_u^{(s)}, n_o[t-1, v_u, j_u(t)], n_C[t-1, v_u, d_u(t-1), M_u\{T_o^{(u)}(t-1)\}, j_u(t)], n_M[t-1, v_u, M_u\{T_o^{(u)}(t-1)\} j_u(t)], n_R[t-1, v_u, d_u(t-1), j_u(t)]; d_u(t'), s_u(t'), j_u(t'), T_o^{(u)}(t'); t' = 0, 1, \ldots, t-1\}$$

$$P_I^{(u)}(d;t) = P_I\Big[d;t, v_u, j_u(t); d_u(t'), s_u(t'), j_u(t'), n_o(t', v, j), n_C(t', v, d', m, j), n_M(t', v, m, j), n_R(t', v, d', j); (t' = 0, \ldots, t-1; j = 1, \ldots, J; v = 1, \ldots, V; d' = 0, 1, \ldots, D; m = 0, 1, \ldots, M)\Big]$$

$$d_u(t) = d\{u, d_u(t-1), s_u(t-1), d_N^{(u)}(t), d_I^{(u)}(t), t_r^{(u)}(t-1), T_R[t, v_u, d_u(t-1), j_u(t)]\}$$

$$T_o^{(u)}(t) = T_o\Big[t, v_u, T_o^{(u)}(t-1), s_u(t-1), d_u(t), d_u(t-1), j_u(t)\Big]$$

$$M_u[T'] = M(u, T', m_u^{(s)})$$

$$t_C^{(u)}(t) = t_C\Big[t, v_u, t_C^{(u)}(t-1), s_u(t-1), d_u(t), d_u(t-1), M_u\{T_o^{(u)}(t)\}, M_u\{T_o^{(u)}(t-1)\}, j_u(t), C_u^{(o)}(t), \tau_u^{(o)}(t), \tau_u^{(M)}(t), \tau_u^{(R)}(t), t_u^{(o)}(t), t_u^{(M)}(t), t_u^{(R)}(t), S_o\{\tau_u^{(o)}(t), v_u, j_u[\tau_u^{(o)}(t)]\}, S_M\{\tau_u^{(M)}(t), v_u, M_u[T_o^{(u)}(t)], j_u[\tau_u^{(M)}(t)]\}, S_R\{\tau_u^{(R)}(t), v_u, d_u[\tau_u^{(R)}(t)], j_u[\tau_u^{(R)}(t)]\}, N_C\{t-1, t_C^{(u)}(t-1), g_C[v_u, d_u(t)\}, M_u\{T_o^{(u)}(t)\}, j_u(t)], j_u(t)\}, N\{t-1, t_C^{(u)}(t-1), g_C[v_u, d_u(t)\}, M_u\{T_o^{(u)}(t)\}, j_u(t)], j_u(t)\}\Big]$$

$$t_m^{(u)}(t) = 0 \text{ if } t_C^{(u)}(t) > 0 \text{ or if } M_u[T_o^{(u)}(t)] = 0 = t_m\Big[t, v_u, t_m^{(u)}(t-1), s_u(t-1), M_u\{T_o^{(u)}(t)\}, j_u\{\tau_m^{(u)}(t)\}, N_M\{t-1, t_m^{(u)}(t-1), g_M(v_u, M_u[T_o^{(u)}(t)], j_u[\tau_m^{(u)}(t)]), j_u[\tau_m^{(u)}(t)]\}, N\{t-1, t_m^{(u)}(t-1), g_M(v_u, M_u[T_o^{(u)}(t)], j_u[\tau_m^{(u)}(t)]), j_u[\tau_m^{(u)}(t)]\}\Big] \text{ otherwise}$$

$$t_r^{(u)}(t) = 0 \text{ if } t_C^{(u)}(t) > 0 \text{ or if } d_u(t) = 0 = t_r\Big[t, v_u, t_r^{(u)}(t-1), s_u(t-1, d_u(t)), j_u\{\tau_r^{(u)}(t)\}, N_R\{t-1, t_r^{(u)}(t-1), g_R(v_u, d_u(t), j_u[\tau_r^{(u)}(t)]), j_u[\tau_r^{(u)}(t)]\}, N\{t-1, t_r^{(u)}(t-1), g_R(v_u, d_u(t), j_u[\tau_r^{(u)}(t)]), j_u[\tau_r^{(u)}(t)]\}\Big] \text{ otherwise}$$

$$t_o^{(u)}(t) = 0 \text{ when } d_u(t) = 0 \text{ and } M_u[T_o^{(u)}(t)] = 0 = t_C^{(u)}(t) + T_M\{t + t_C^{(u)}(t), v_u, M_u[T_o^{(u)}(t)], j_u^{(M)}[t + t_C^{(u)}(t)]\} \text{ when } M_u[T_o^{(u)}(t)] > 0 \text{ and } t_C^{(u)}(t) > 0 = t_C^{(u)}(t) + T_R\{t + t_C^{(u)}(t), v_u, d_u(t), j_u^{(R)}[t + t_C^{(u)}(t)]\} \text{ when } d_u(t) > 0 \text{ and } t_C^{(u)}(t) > 0 = t_m^{(u)}(t) \text{ when } M_u[T_o^{(u)}(t)] > 0 \text{ and } t_C^{(u)}(t) = 0 = t_r^{(u)}(t) \text{ when } d_u(t) > 0 \text{ and } t_C^{(u)}(t) = 0$$

$$j_m^{(u)}(t) = 0 \text{ when } M_u[T_o^{(u)}(t)] = 0 \text{ or } t_C^{(u)}(t) > 0 = j_u[\tau_m^{(u)}(t)], \text{ otherwise}$$

$$j_r^{(u)}(t) = 0 \text{ when } d_u(t) = 0 \text{ or } t_C^{(u)}(t) > 0 = j_u[\tau_r^{(u)}(t)], \text{ otherwise}$$

$$s_u(t) = 1 \text{ when } t_o^{(u)}(t) = 0, t_C^{(u)}(t) = 0$$
$$= 2 \text{ when } t_o^{(u)}(t) = 0, t_C^{(u)}(t) > 0$$
$$= 3 \text{ when } t_o^{(u)}(t) > 0, t_C^{(u)}(t) > 0, M_u[T_o^{(u)}(t)] > 0$$
$$= 4 \text{ when } t_o^{(u)}(t) > 0, t_C^{(u)}(t) > 0, d_u(t) > d_u(t-1)$$
$$= 5 \text{ when } t_o^{(u)}(t) > 0, t_C^{(u)}(t) > 0, d_u(t) > 0, d_u(t) = d_u(t-1)$$
$$= 6 \text{ when } t_o^{(u)}(t) > 0, t_C^{(u)}(t) = 0, M_u[T_o^{(u)}(t)] > 0$$
$$= 7 \text{ when } t_o^{(u)}(t) > 0, t_C^{(u)}(t) = 0, d_u(t) > 0$$

$$q_u[t, v_u, j_u(t), j_2] = q\{t, u, v_u, j_u(t), j_2, s_u(t-1), d_u(t), T_o^{(u)}(t); T_o^{(u')}(t), m_s^{(u')}, \text{ for all } u', \text{ such that } v_{u'} = v_u, j_{u'}(t) = j_u(t), s_{u'}(t-1) = 1, d_{u'}(t) = 0 \text{ and } M_{u'}[T_o^{(u')}(t)] = 0\}$$

$$n_1(t, v, j) = n_1\{t, v, j, n_o(t-1, v, j); v_u, d_u(t), M_u[T_u^{(o)}(t)], \text{ for all } u, \text{ such that } v_u = v, j_u(t) = j, \text{ and } s_u(t-1) = 1\}$$

$$n_a(t, v, j_1, j_2) = \min\{n_s(t, v, j_1, j_2), \max[0, n_1(t, v, j_1) - B_v(t, j_1)]\}$$

$e(t)$ = function that can depend on any of the simulated actual results (up to end of t-th interval) and any of the anticipated future results for all possible combined choices for the present and future strategies.

$e[t';t, t_e(t)]$ = function that can depend on any of the actual results (up to end of t-th interval) and any of the anticipated future results that are obtained on basis of situation at end of t'-th interval, where $t' \leq t + t_e(t)$.

$n_S(t, v, j_1, j_2), S_O(t, v, j), S_M(t, v, m, j), S_R(t, v, d, j)$

= strategies that are chosen from all combinations of possibilities by anticipating future behavior of system, along lines specified in the Definition of Notation, and use of $e(t)$; determines $J_O(t, v, j)$.

$N_C(t, t', g_C, j), N_M(t, t', g_M, j), N_R(t, t', g_R, j), N(t, t', g, j)$

= functions that are directly obtained by anticipating future behavior of system (as outlined in the Definition of Notation) using the selected strategies.

This completes the statement of the functional relationships among the various quantities that occur in the simulation model presented. The mathematical model furnished by these relationships, and their order of presentation, seems to be sufficiently general for use in the investigation of many types of logistics situations.

* * *

PANEL 3b
MICROSIMULATION IN LOGISTICS RESEARCH

A COMPARATIVE STUDY OF PREDICTION TECHNIQUES*

Bernice Brown

The RAND Corporation
Santa Monica, California

This study deals with the techniques for predicting the future demands of spare parts. The results of studies about the Air Force Inventory problem have led us to believe that any improvement in the demand forecast could result in important savings. Demand prediction is actually based upon two separate estimating procedures. First, an estimate is made about the underlying pattern of demands per program element. Second, an estimate is made about the pattern of future program elements. It is the marriage of these two estimates that allows us to make predictions of the number of demands in future time periods. This study is restricted to finding ways of improving the estimates of the demands per program element. We used past demand data with the pattern of program elements known. We also used the known pattern of future program elements to make the prediction of future demand.

There are seven prediction techniques used in this study. The first four techniques are based on the assumption that demand rate (demand per program element) is constant through time. The other three techniques assume that failures (demands) are related to the age of the part. These three techniques are routines for estimating the service life characteristics of parts.

We used as inputs to the study two batches of data: 33 months of demand-data on a sample of 300 B-52 parts and 26 months of failures of 27 missile components. We made monthly estimates of demand for a year in the future and measured the prediction errors. In the evaluation of the estimating procedures, we used three measures of the accuracy of the forecasts, namely average monthly error, relative error expressed as percent of the actual demand and root mean square of the errors of the predictions.

SUMMARY AND CONCLUSIONS

The evaluations made in this study are comparative rather than absolute. As a result, all we can say is that one technique yielded smaller errors than did another, since even the preferred technique for a particular part will show some errors of prediction. We make errors when we accept the observed demand over a specified period as the mean of the distribution of demands. We also make errors when we use the estimated mean demand to predict the number of demands in the future period. We may make errors when we assume

*Prepared as RAND Corporation paper P-1980.

that the relationships between past and future demands are a function only of the total program activity in each period. We may make errors when we assume that we can measure a part's age by the number of flying-hours logged by the aircraft on which it is installed.

A number of observations may be put forth as tentative conclusions:

(1) About 25 percent of the sample parts showed zero demand through 21 months of data. These are considered "unpredictable" by any of these statistical techniques.

(2) Issue-interval Technique 1 underestimated much of the time, especially in the early months of the program.

(3) Technique 1 appears to be more accurate than the other techniques for engine parts and communications (Groups 1 and 5).

(4) Technique 6 is preferred among the service-life techniques, but fewer than 10 percent of the sample line items show evidence of age-related demands.

(5) One of the upper-bound techniques (4) seems to be preferred for gunnery and accessories (Groups 3 and 6) with a switch to Technique 1 if the demand rate stabilizes.

(6) Upper-bound Technique 2 seems adequate for airframes (Group 2).

(7) Upper-bound Technique 3 seems preferred for fire-control components (Group 4).

A great deal of work is yet to be done. These seven prediction techniques were applied to data on failures during the checkout of 27 components of an air-to-air missile, and results of that study need to be synthesized with the results reported here. A study has also been made which uses the engineering estimate of a service-life factor delineated early in the program, before any demand data have been collected. One solution to the demand-prediction problem may involve a routine which originates with the first projection of an engineering estimate and systematically uses two or three of these techniques, weighted in such a way as to take advantage of evolving information about part changes, program changes, maintenance policy changes, etc. We do not expect any of these studies to be the last word on demand-prediction techniques.

INTRODUCTION

Predicting the future demands of aircraft spare parts in an important step in managing the Air Force inventory. Many studies of the Air Force inventory problem point out one salient fact: an improvement in demand prediction could result in important savings, perhaps more important than could any other single change; even a small improvement is likely to yield significant savings. This study discusses some promising methods for improving predictive techniques.

Demand prediction is based on two separate estimating procedures. First, an estimate is made of the underlying pattern of demands per program element, and second, the pattern of future program elements is estimated. The marriage of these two estimates allows us to predict the number of demands in future time periods; an improvement in either estimate would accordingly improve predictive accuracy. This study is restricted to finding ways of improving the first estimate. We are using past demand data with the pattern of program elements known. We will use the known pattern of future program elements in making our predictions about future demands. This will allow us to concentrate on our primary aim of finding ways of improving the estimates of demands per program element.

In the Air Force supply system, we have long known that a knowledge of the future demand for aircraft spare parts is necessary for effective and economical procurement, distribution, and stockage decisions. Many people have attempted to reduce, and thus control the

uncertainties of the future, by making estimates of future demands using various techniques to measure their accuracy. The precision of his estimates reflects the forecaster's mastery over uncertainty.

The methods compared in this study have been designated as Prediction Techniques 1 through 7. It will be convenient to describe each one briefly.

Technique 1

We divide the total number of demands in the experience period by the number of flying-hours in the period. The rate thus obtained is multiplied by the planned number of flying-hours for the prediction period to obtain the predicted number of demands for that period:

(1) $$x_2 = x_1 \frac{T_2}{T_1},$$

where

x_1 = number of demands during experience period;
x_2 = number of demands predicted;
T_1 = number of flying-hours in the experience period; and
T_2 = number of flying-hours in the prediction period.

Technique 2

The observed demands during the experience period are used to enter a table giving the upper 90-percent confidence limit for the mean of a Poisson variable [1].

(2) $$x_2 = x_1^{(ul)} \frac{T_2}{T_1}.$$

The upper confidence limit is then multiplied by the ratio of the number of flying-hours in the prediction period to the number of hours in the experience period to obtain the predicted number of demands.

Technique 3

This is a nonparametric Poisson technique using the following equation:

(3) $$(p\,x_2 - q\,x_1)^2 = p\,q\,k^2\,(x_1 + x_2),$$

where:

k = 1.65 (value of normal deviate for 90-percent confidence interval), and

$$p = \frac{T_1}{T_1 + T_2} = 1 - q.$$

Technique 4

This technique is based on the following equation:

$$x_2 = \left[(x_1 + 1) + 1.65\sqrt{3\ (x_1 + 1)}\right]\frac{T_2}{T_1}\ , \tag{4}$$

where the symbols have the same interpretation as in Technique 3.

Technique 5

This technique is an adaptation of the Minimum Normit Chi-Square method as described by Berkson [2]. The prediction is the value which minimizes the following sum of squares:

$$\text{Min}\Sigma n_i\ w_i\ (\hat{t}_i - t)^2\ , \tag{5}$$

where

Σn_i = total number of exposures;
$1/\ n_i w_i$ = estimate of variance of $\hat{t}_i$;
$\hat{t}_i = a + b\,x$:
x = age at time of demand; and
a and b are parameters of the regression of the normal deviate t on age x.

Technique 6

This technique differs from 5 only in that we use the transformation log (1 + age) instead of using the raw age data. The sum of squares to be minimized may be written as follows:

$$\text{Min}\Sigma n_i\ w_i\ (\hat{t}_i - t)^2\ , \tag{6}$$

where
Σn_i and $1/n_i\ w_i$ are the same as in 5 above;
$\hat{t}_i = a + b\left[\log\ (1 + x)\right]$;
x - age at time of demand; and
a and b are parameters of the regression of the normal deviate t on log (1 + x).

Technique 7

This technique is the actuarial method used by the Air Force to predict engine failures. It is described in T.O. 00-25-128, 18 April 1958. The forecast of failures is made by use of a standard mortality table which contains the percentage of survivors at the beginning of each age interval and the percentage failing in the interval. The mortality table is developed by using the data on failures and exposures for each age interval, smoothing the crude failure rates, and fitting a curve to the smoothed rates. The predictions using this technique were made at Hq AMC under the direction of Mr. John L. Madden, MCMSM. Max Astrachan, RAND consultant, coordinated the work on this technique.

In Techniques 1, 2, 3, and 4, the assumption is made that demand rate (number of demands per thousand flying-hours) is constant through time.

In Techniques 5, 6, and 7, the fundamental assumption is that demand is related to the age of the part. These three techniques may therefore be considered as routines for estimating the service-life characteristics of parts.

For convenience, we have called Technique 1 "Issue Interval," Technique 2 "Upper-bound Poisson Mean," Technique 3 "Upper-bound Non-parametric Poisson," Technique 4 "Upper-bound Normal," Technique 5 "Regression on Age," Technique 6 "Regression on Log" (1 plus age), and Technique 7 "Actuarial."

DATA USED

The data on B-52 parts consumption came from the Irvine reporting system. We had complete reports of transactions from two bases - Loring AFB and Castle AFB—for a period of 33 months from January 1956 through September 1958. We chose to study Hi-Valu and Category II parts only, in six major property classes. These six classes were: engine components, airframe structural components, gunnery components, bombing fire control components, communications equipment, and aircraft accessories. About 7500 part numbers and 46,000 units of demand were listed for these six property classes. (Demand as used here is the sum of Transaction Codes 1, 5, and 9 on the consumption-data cards.)

About a thousand part numbers showed demands of five units or more in the 33-month period at the two bases. We researched these part numbers, matching AF part numbers with Federal stock numbers. We tried to get information on substitute and interchangeable parts. We used the 231 forms from the Hi-Valu program and catalog information to amalgamate part numbers that belonged together. We used information on families of parts furnished us by OCAMA, as far as it went. Combinations and deletions gave us 875 parts with five demands or more during the 33-month period.

The distribution of these parts by property class is as follows:

	Property Class	Code	No. of Parts With Demands of 5 or More (Jan 56 - Sept 58)
1.	Engine Components	02H1	139
2.	Airframe Structural Components	1AFG	166
3.	Gunnery Components	Fed 1270	78
4.	Bombing Fire Control Components	Fed 1280	113
5.	Communications Equipment	Fed 5800	136
6.	Aircraft Accessories	AF 03B, C, F, I	243
		Total	875

We chose 300 parts from the listing of 875 parts, taking 50 from each property class. We drew random numbers for the designation of 50 sample parts from each group, using the ordering of the part numbers as they appeared in the original machine listing. We do not consider these six property classes as representative of the population of all B-52 spare parts. However, the parts used in the study are a random sample of parts with demands of five or more in each of the designated property classes.

About 40 percent of the parts in the sample were Hi-Valu parts and 60 percent were Category II reparable parts.

Deletions and combinations of parts reduced the original sample to 272 parts, to which the first four prediction techniques were applied. The criterion for using Techniques 5 and 6

was that no prediction could be made until at least four demands had been experienced, which limited their application to 145 parts. To use Technique 7 (the actuarial method) it was desirable to have accumulated about 50 demands for the part. We modified the requirement by stating that 10 or more demands must be experienced in the first year, in order to make a prediction at the end of the 12th month. The technique was thus used on only 12 parts. It was also necessary to restrict its use to parts with a single application on the aircraft.

The following table shows the number of parts in each property class to which each technique was applied:

TABLE 1

Property Class	Technique No.		
	1, 2, 3, 4	5, 6	7
	No. of Parts		
1. Engine	47	9	-
2. Airframe	45	35	1
3. Gunnery	46	30	-
4. Fire Control	38	24	4
5. Communications	48	16	5
6. Accessories	48	31	2
Total	272	145	12

During the 33-month period covered by the data, about 68,000 flying-hours were accomplished. The program included 2,500 aircraft-months, with an average flying program close to 30 hours per plane-month. The flying-hour program built up slowly: about 55 percent of the total flying-hours had been accomplished by the end of the 21st month, with 45 percent of the flying being accomplished in the last year of the data (months 22 through 33). About 42 percent of the demands for the 272 sample parts occurred during the first 21 months, about 58 percent during the last year.

COMPARISON OF TECHNIQUES FOR ACCURACY OF ESTIMATES WHEN MONTHLY PREDICTIONS ARE MADE FOR A YEAR IN THE FUTURE

We used three measures of accuracy in evaluating the predictions: the average monthly error, the relative error expressed as a percentage, and the root mean square of the errors. These three measures are defined as follows:

$$\text{Average monthly error} = \frac{\Sigma(E - A)}{N};$$

$$\text{Relative error} = \frac{\Sigma(E - A)}{\Sigma A}; \text{ and}$$

$$\text{Root mean square} = \sqrt{\frac{\Sigma (E - A)^2}{N}},$$

where:

E is estimated number of demands in a given month;
A is observed number of demands in a given month; and
N is number of months.

Using data on demand cumulated to the end of each month, we made monthly predictions of demands for the ensuing year for each of the applicable techniques. The number of

flying-hours was used as the program element for projecting the forecast. We had 21 sets of predictions for Techniques 1, 2, 3, and 4, substantially fewer than 21 sets for Techniques 5 and 6, and only four for Technique 7.

There were 70 parts in the sample that had no demands in the first 21 months. It is not necessary to look at prediction errors in order to compare techniques for this group of parts, since the prediction is simply a constant multiplied by the ratio of future to past program experience. For Technique 1, zero demand experience gives zero demand prediction, regardless of the program element. For Technique 2, zero experience gives 3 (T_2/ T_1) as the prediction. For Technique 3, zero experience gives 2.7225 (T_2/ T_1) as the prediction and for Technique 4, 3.8578 (T_2/ T_1).

To repeat, low-average demand is characteristic of many aircraft spare parts. For about one-fourth of our sample B-52 parts, we have only zero demand experience. The distribution of these 70 parts in the six property classes reveals that 35 are among engine components.

Table 2 shows for the 12-month period from October 1957 through September 1958 (months 22 through 33 in the data), the demand history of this group of 70 parts which had zero demand in the first 21 months.

It is clear that none of our prediction techniques will be useful in making monthly forecasts for these parts. The predictions will be adequate only for the 474 distribution-months in which no demands occurred. For month-by-month predictions, the ratio of flying-hours in the prediction period to flying-hours in the experience period will be less than one-tenth. The question arises: Can we make a satisfactory prediction for the 12-month period? The flying-hour ratio for the year is about 0.82. Technique 4 thus gives us a predicted value of about 3 units, which will cover the annual demand of only 2 of the 70 parts. Demands for these parts range from 1 to 113 units; all but 2 will be underestimated by the statistical techniques described in this study. Approximately 40 percent of these estimates will show relative errors of less than -50 percent. The part which showed a demand of 113 units—Engine Component, Vane and Shroud Assembly (8 in each aircraft)—will be underestimated by 110 units, showing a relative error of - 97 percent. For this group of parts, it seemed fruitless to look at measures of accuracy.

Predictions were made for 145 parts, using six techniques. Technique 7 was used on 12; on 57, only Techniques 1 through 4 could be used. We used the average monthly error of predictions, made for a year in the future, to evaluate the accuracy of the predictions. Using this measure of accuracy as a criterion, we can compare the techniques and pick a preferred technique for each part. We examined these parts by property class groups.

Engine Components

The average monthly error for 9 parts is shown in Table A-1 in the Appendix. For 6 of the 9, Technique 1 appears to give the smallest average monthly error, and for 3 of the 9, Technique 3 is preferred as giving the smallest error. For 3 parts, only four techniques were used, of which Technique 1 proved the most accurate. It was possible to use service-life Techniques 5 and 6 on only 9 of the 47 sample parts in this property class. None had enough demands to use Technique 7.

The salient fact about the sample parts in this property class is that three-fourths of them had zero demands through the first 21 months of the program. Zero demands beset even

the 9 parts which had enough demand to use all six techniques: in 33 months' experience with these 9 parts, there was no demand at all in about half. Probably no technique will enable monthly predictions for these parts. After 21 months of experience it is possible to make predictions for a year in advance for 12 parts, with relative errors ranging from -80 to +130 percent on 10 of them. For two parts (109 and 111), the demand is zero in the last year of the data record, so that all predictions are gross overestimates.

TABLE 2

Demands For Each of 12 Months Following a History of 21 Months of Zero Demand
(No. of Parts by Property Class and Frequency of Distribution
of Monthly Demand of 70 Parts for 12 Months)

Month	Engine	Air-Frame	Gunnery	Fire Control	Communi-cations	Acces-sories	Total	Cumu-lative Percent
0	226	11	28	38	127	44	474	56
1	43	--	15	23	16	11	108	69
2	47	--	11	7	8	3	76	78
3	22	--	2	5	10	3	42	83
4	18	--	3	4	7	2	34	87
5	7	--	1	3	4	2	17	89
6	15	1	--	2	4	3	25	92
7	9	--	--	1	2	1	13	94
8	11	--	--	1	1	--	13	95
9	5	--	--	--	2	--	7	96
10	5	--	--	--	2	--	7	97
11 up	12	--	--	--	9	3	24	100
Totals	420	12	60	84	192	72	840	--
Total Demand	893	6	60	110	378	113	1560	--
No. of Separate Parts	35	1	5	7	16	6	70	--

Airframe Structural Components

The average monthly error of predictions, made for a year in advance, is shown in Appendix Table A-2 for the 35 parts to which six techniques were applied. Actuarial Technique 7 was applied to only one part: Part #222, (Door Assembly), shown in row 18 of Table A-2.

We examined the table to pick a preferred technique on the basis of average monthly error, with the following results: Sometimes it was necessary to look at relative errors in order to choose between equally good or equally poor techniques. We chose Technique 1 for 19 parts, and switched to Technique 6 for one of them. We chose Technique 2 for 12, and switched to Technique 6 for two. We chose Technique 3 for 4, and switched to Technique 6

for two. We found 5 sample parts in the airframe group for which Technique 6 is preferred as soon as enough information is available so that the technique can be used to make predictions. (These five parts are 206, 219, 242, 243, and 247—two struts, an actuator, cover assembly, and cartridge assembly.) For 9 parts, only the first four techniques were used. Among these parts, 5 give smaller errors with Technique 1 and four with Technique 2. One part in the airframe class was observed to have no demands in the first 21 months of the data.

Summarizing over the 44 sample parts, we have Technique 6, a service-life routine based on the assumption that demand is a function of log of one plus age, for 5 parts (11 percent); Technique 3, an upper-bound technique based on Poisson assumption for 2 parts (5 percent); Technique 2, an upper-bound on the Poisson mean, for 14 parts (32 percent), and Technique 1, the projection of a constant mean demand rate, for 23 parts (52 percent).

If we used the preferred technique indicated above and made a prediction for a year in the future at the end of 12 months' experience, we have negative relative errors on 14 of the 39 parts (5 parts either zero experience or zero in prediction period). If we do so at the end of the 21st month, we have negative relative errors on 7 of the 40 parts (4 zero in the months 22 - 33). The magnitude of the relative errors for the year's prediction varies from -92 to +600 made at the end of the 12th month; and from -72 to +700 for the prediction made at the end of the 21st month.

Gunnery Equipment

Appendix Table A-3 shows the average monthly error of predictions made for the year immediately following the experience period, using Techniques 1 through 6. The table shows 30 parts with four different experience periods. For 5 parts there were demands in the first 21 months. For 11 more, only Techniques 1 through 4 were used; of these, 5 gave smaller prediction errors with Technique 1, 3 with 3 and 3 with 4.

In Table A-3, 3 parts so strongly appear to have age-related demands that we would prefer Technique 5 for parts 301 and 314, and Technique 6 for 304. However, the reader will note that with Part 301, 18 months elapse before there is enough information to use Technique 5; and 15 months have so elapsed for Parts 304 and 314. It would therefore be necessary to use one of the first four techniques until enough experience is available. Since the prediction errors are small for all four techniques in the first 12 to 18 months, we would recommend using one of upper-bound techniques in the early period; we would choose Technique 4 as giving the greatest upper bound. For the other 27 parts, Technique 1 is preferred for 9; Technique 3 for 7; and Technique 4 for 11.

In trying to evaluate the numbers in Table A-3, it often was necessary to look at the relative error and root mean square in order to choose a single preferred technique. With Part 344, for example, there is no choice between Techniques 1 and 6 in regard to accuracy as measured by average monthly error, but we chose Technique 1 when both relative error and root mean square were considered in the evaluation.

Only 4 parts among the 30 show average monthly errors greater than 5 units, and about 12 show average monthly errors of less than one unit. For the 30 parts in Table A-3, 75 percent of the total number of units of demand occurs in the last year of the data. The effect of this is shown in the consistency with which all the techniques underestimate the actual demands for about half the parts in the table. There seems to be more homogeneity in the behavior of the parts in this class than in either the engine or airframe groups. But

the consistent underprediction indicates that the assumption of a constant demand-rate through time has been violated. There is some indication that something happened about the end of the 12th month to stimulate an increasing demand rate.

If we had to choose a preferred technique for the parts in this property class, it would be Technique 4. But at the end of 12 months for Parts 301, 310, 314, 320, 321, 322, 323, 327, 331, 334, 335, 338, 343, the rule we would prescribe would be to give up the data on demand and program element for the first year and make new estimates of the demand rate.

Bombing Fire Control Components

The average monthly error of predictions made for a year in the future is shown in Appendix Table A-4 for the 24 parts for which 6 techniques were used. Actuarial Technique 7 was applied to 4 parts: 418 control, 421 amplifier, 433 power supply, and 443 amplifier. For the Computer Control Box, Technique 6 is preferred over 7, but 7 is preferable for the other three parts. However, some other prediction technique must necessarily be used during the first year of the data.

We examined the table to pick a preferred technique on the basis of average monthly error, with the following results. We chose Technique 1 for 9 parts, with a later switch to Technique 4 for two parts and to Technique 7 for one. We chose Technique 2 for 6, with a switch to Technique 6 for one. We chose Technique 3 for 7, with a switch to Technique 1 for one, to Technique 6 for one, and to Technique 7 for two. We chose Technique 4 for two parts.

There were 7 parts for which only the first four techniques were used. Among these parts, 3 give smaller average monthly errors with Technique 1, 3 with Technique 3, and 1 with Technique 4. There were 7 parts in this group of electronic components for which no demands were observed during the first 21 months of the data collection.

Summarizing over the 31 sample parts, we find service-life Techniques 6 and 7 applied for 5 parts (16 percent); upper-bound Techniques 2, 3, and 4 preferred for 15 parts (55 percent); and issue-interval Technique 1 preferred for 11 parts (29 percent).

Communications Equipment

We looked at average monthly error of predictions made at the end of the 12th, 15th, 18th, and 21st months for a year in advance. The errors are shown in Appendix Table A-5 for 16 parts in this property class. The sample parts are from Air Force Classes 16 A, K_1 and K_2, or Federal Class 5800. There were 48 parts in the B-52 sample, of which 16 showed no demand during the first 21 months. Predictions were made for 32 by using Techniques 1 through 4 only, but we will not evaluate the results for the 16 parts for which predictions were made with zero demand information. In 10 cases, the preferred routine would be to use Technique 1 during the first 10 to 12 months or until 4 or 5 failures had been experienced, and then use Technique 3. In 6 cases, the preferred routine was to use Technique 1 during the first 10 to 12 months and then use Technique 4.

Techniques 1 through 6 were used to forecast demands for 16 parts, of which 5 were used as inputs for the application of actuarial Technique 7.

The following evaluation was made from Table A-6:

Part 503 - 1 shows smallest errors.

506 - 1 and 7 very close in first 2 periods—7 is better than 5 or 6—look at relative error and choose 1 over 7.

507 - 5 and 6 both better than 1, 2, 3, 4—look at relative errors and choose 6.
508 - 6 and 7 look better in the first two periods but several of the other techniques seem all right later on. Look at relative error and choose 7.
513 - Errors small on 3 and 4—look at relative error and choose 4.
515 - Very mixed up. Look at relative error and choose 6 over 1.
516 - 6 considerably better than 5 but 1 gives small errors also. Look at relative error and choose 1.
518 - 6 better than 5 but 1 and 3 give even smaller errors. Look at relative error and choose 3.
519 - Choose 6.
523 - Very low demands (total 7). Technique 1 suitable.
524 - Choose 6 over 1.
528 - Errors small all around; look at relative error and choose 4.
544 - 1 and 6 close—look at relative error and choose 6, but it is still close.
546 - 1 and 6 close—look at relative error and choose 1.
548 - 5, 6, and 7 all show improvement over first four techniques; choose 6.
550 - 6 and 1 close; look at relative error and choose 6.

The summary of the choices for these 16 parts is: Technique 1 for 5 parts (receiver, transmitter, dynamotor, power supply, tube); Technique 3 for 1 (control for radar set); Technique 4 for 2 amplifiers; Technique 7 for 1 receiver-transmitter; and Technique 6 for 7 parts (2 tubes, converter, indicator, relay, motor for chaff dispenser, and control). Since the service-life techniques can not be used until some demand data have accumulated, one of the first four techniques must be used early in the period. The preferred technique for the 32 parts in this property class may be summed up as follows: Use Technique 1 on 24 parts, switching to 3 on ten, to 7 on one, to 4 on six, and to 6 on two. Use Technique 2 in the early stages on three parts, with a switch to Technique 6. Use Technique 3 on one part. Use Technique 4 on four parts, switching to Technique 6 on two.

Aircraft Accessories

This group of parts comprises Air Force property subclasses 03B, 03C, 03F, and 03I. Federal Supply Classification Groups include 1630, Aircraft Wheels and Brakes; 1650, Aircraft Hydraulic Systems; 1660, Aircraft Air-Conditioning and Heating Equipment; 1680, Miscellaneous Accessories; 2915, Fuel System Components; 6105, Electrical Motors; and 6115, Generators. There were 48 parts in the B-52 sample, of which 6 showed no demand during the first 21 months. For 11, only Techniques 1 through 4 were applied. For 31, Techniques 1 through 6 were used. Technique 7 was used for two.

We looked at the average monthly error of predictions made at the end of each month for a year in the future. Appendix Table A-6 shows the errors or predictions made at the end of the 12th, 15th, 18th, and 21st months. The preferred technique picked on the basis of the errors shown in Table A-6 may be stated as follows: Technique 1 is preferred for 15 parts, with a change to Technique 6 indicated on one part. Technique 2 is preferred for two parts. Technique 3 is preferred for eight parts, with a change to Technique 6 on two parts. Technique 4 is preferred for six parts, with a change to Technique 6 on Part No. 646. For the 11 parts for which predictions were made with Techniques 1, 2, 3, and 4 only, Technique 1 is preferred for eight and Technique 4 for three.

The preferred techniques for the 42 parts may be outlined as follows:

(1) Technique 1 is preferred for 23 parts, with a change to 6 on one part.

(2) Technique 2 is preferred for 2 parts.

(3) Technique 3 is preferred for 8 parts, with a change to 6 on 2 parts.

(4) Technique 4 is preferred for 9 parts, with a change to 6 on one part.

Summarizing over the parts in this property class, Technique 1 seems to give greater all-around accuracy; Techniques 3 and 4 are about equally preferable as upper-bound techniques; and Technique 6 is preferable to 5 and 7 as a service-life technique.

Summary Of Preferred Techniques For Sample Parts

We have looked at the sets of predictions made for each part and compared techniques on the basis of the accuracy of the estimates. For several parts, the preferred routine was to use one technique for the first year and then change to a second technique. In order to present a summary, two decisions were made: (1) When one of the service-life techniques was indicated for a part, we have allocated the part to that technique even though we know it will not be possible to use it until enough data are available; and (2) When the preferred routine indicated switching among the first four techniques, we did not permit the change but listed the part under the first technique preferred. The summary is presented in Table 3.

TABLE 3

Technique Preferred as Giving Smallest Prediction Errors
(Summary of 202 Parts, by 6 Property Classes)

Property Class Group	Assume Demand-Rate is Constant				Assume Demand is Related to Age			Total No. of Parts for Each Property Class
	Issue-Interval Technique	Upper-Bound on Issue Interval Technique			Service-Life Techniques			
	1	2	3	4	5	6	7	
Engine	9	--	3	--	--	--	--	12
Airframe	23	14	2	--	--	5	--	44
Gunnery	14	--	10	14	2	1	--	41
Fire Control	11	5	7	3	--	2	3	31
Communications	21	2	1	2	--	7	1	32
Accessories	22	--	6	8	--	4	--	42
Totals	100	21	29	27	2	19	4	202

It will be seen in the Totals that service-life techniques are preferred for 25 parts, the issue-interval technique for 100, and the three upper-bound techniques for 77. Let us recall that we had a sample of 272 B-52 parts with demand and program data which constituted the inputs for the study. For 70 parts (about 26 percent) there was zero demand for the entire period. This group is not included in Table 3. We consider these parts "unpredictable" and

did not examine the prediction errors. This group is further examined below under "Discussion of Results." Technique 1 is preferred for 100 parts (about 37 percent). The three upper-bound Techniques 2, 3, and 4 account for 77 parts (about 28 percent), with nearly equal representation among the three. Technique 2 is the least upper-bound, 4 is the greatest, and 3 is intermediary. Techniques 5 and 6 are preferred for 21 parts (about 8 percent), with 6 greatly favored over 5. Technique 7 was preferred for 4 parts (a little more than 1 percent).

It is possible to generalize somewhat and state that Technique 1 would be satisfactory across the board as the representative of the constant-demand-rate techniques for the engine parts and the communications parts. Since the upper-bound techniques are close together, 1 and 4 could be used for gunnery and accessories, 1 and 2 for airframe, and 1 and 3 for fire control.

DISCUSSION OF RESULTS

Factors other than predictive accuracy must be considered. It is well known that we regard an underestimate as more serious than an overestimate, but even the experts will disagree about the factors. However, in this study we have made monthly predictions for the future on 272 parts. We made 252 monthly projections for each technique. But not all techniques were applied to all parts. The data on the 145 parts to which Techniques 1 through 6 were applied furnish us with a count of the number of monthly predictions which were underestimates of the observed demand, and of the number which were overestimates. We can standardize these numbers and assign weights to each of the techniques; the sum of the weighting factors will be unity. Similarly, we can use the probabilities obtained from Table 3 as weighting factors for ranking the techniques according to their predictive accuracy. By the same method, we can inspect the tables in the Appendix (and measures of relative error and root mean square not included in the report) and thus be able to assign weight factors to each technique on its ability to improve in accuracy as time moves on and additional information unfolds. We can get some insight into the relative costs and complexity of the data and of the computation requirements by using the machine-time recorded, the programming-time required, etc., to give us relative numbers.

By using a curious mixture of objective numbers and subjective judgment we have developed a table of weights for five factors which we believe to be important in comparing the prediction techniques. The allocation of weight factors is presented in Table 4.

Techniques 2 and 4 have the same rating and should both be ranked number 1, but both 1 and 3 are very close to 2 and 4 and should be ranked number 2. Technique 6 is clearly the leader of the three service-life techniques. It may be noted that Technique 5, which showed the largest prediction errors, at the same time showed the smallest number of underestimates. Technique 5 did not perform well; it showed a decided tendency to overestimate. The object of trying to combine these factors and rate the various techniques was to be able to say that if we had to choose a single technique to be used across the board, we would choose 2 or 4. At least it appears in this study that the calculated risk of using 2 or 4 is less than for the other techniques. Technique 2 assumed the existence of a Poisson world and 4 a normally-distributed world.

Under "Comparison of Prediction Techniques by Use of Weighting Factors" above, we looked at a series of prediction errors for each technique and each part. For an overall picture, let us look at the history of the 272 parts at a specific point in time: the end of 21 months

TABLE 4

Comparison of Prediction Techniques by Use of Weighting Factors*

Technique	Predictive Accuracy	No. of Underestimates	Improvement in Accuracy	Responsiveness or Flexibility	Cost and Complexity of Data & Computation	Total	Rank
1	0.01	0.29	0.13	0.05	0.04	0.52	4
2	0.05	0.20	0.09	0.09	0.05	0.48	2
3	0.04	0.15	0.14	0.11	0.07	0.51	3
4	0.04	0.18	0.09	0.11	0.05	0.47	1
5	0.49	0.04	0.18	0.19	0.16	1.06	6
6	0.05	0.14	0.14	0.19	0.17	0.69	5
7	0.32	—	0.23	0.26	0.46	1.27	7

*The Low weight factor is associated with good performance; i.e., high accuracy (small prediction errors), small number of underestimates, good ability to improve accuracy, high responsiveness, low cost, and complexity. The sum of the weights is unity for each of the five factors.

of data-collection and program experience. This is the date of the last yearly projection, using the demand-experience period with the maximum amount of demand information.

Table 5 is a two-way table with 12 rows and 12 columns showing the distribution of the 272 sample parts, classified by the amount of demands in the experience period (first 21 months) and the amount of demand in the prediction period (last 12 months). For example, the number "8" in the 2nd row, 4th column, signifies that 8 parts had demands of 1-5 in the first 21 months and 11-15 in the last 12 months. About 55 percent of the total flying-hours had been accomplished by the end of the 21st month, so that we can think of the ratio of T_2 to T_1 as approximately one. Demand in the experience period becomes the estimate of demand in the last year if we use Technique 1. If all 272 parts were on the NW-SE diagonal of the chart, we would consider that issue-interval Technique 1 had provided satisfactory estimates.

The upper-bound Poisson mean, Technique 2, gives an estimate one block above the diagonal, and Technique 4 gives an estimate approximately two blocks above the diagonal; Technique 3 is intermediate between them. The following results were obtained by an inspection of Table 5:

54 parts on the diagonal;
66 parts are below the diagonal;
28 parts one block above the diagonal;
32 parts two blocks above the diagonal.

These numerical results may be interpreted as follows:

Technique 1 would have covered 120 parts (44%);
Technique 2 would have covered 148 parts (54%);
Technique 4 would have covered 180 parts (66%).

TABLE 5

Distribution of 272 Sample Parts, Classified by Demands in The First 21 and in The Last 12 Months

Demands in First 21 Mo.	Demands in the Last 12 Months												
	0	1-5	6-10	11-15	16-20	21-25	26-30	31-35	36-40	41-45	46-50	51 up	Total
0	-	5	18	13	12	3	2	4	3	1	3	6	70
1-5	3	23	15	8	4	3	4	1	1	1	-	3	66
6-10	7	11	7	3	2	1	-	1	1	-	-	3	36
11-15	1	8	3	-	1	-	1	-	1	-	-	2	17
16-20	-	1	2	3	4	-	1	-	3	-	-	4	18
21-25	-	2	1	-	1	-	-	1	-	-	-	2	7
26-30	-	-	-	-	1	-	-	3	-	-	-	6	10
31-35	1	-	-	-	1	-	-	-	-	-	-	1	3
36-40	1	1	1	-	-	-	-	-	1	-	-	2	6
41-45	-	1	-	-	-	1	1	-	-	1	-	2	6
46-50	-	-	1	1	-	-	1	-	-	-	-	1	4
51 up	-	-	-	2	1	1	1	3	2	1	-	18	29
Total	13	52	48	30	27	9	10	14	12	4	3	50	272

If we omit the 70 parts for which there was no demand in the first 21 months and base the results on 202 parts, the percentage of parts covered by Techniques 1, 2, and 4 would increase slightly:

Technique 1 would have covered 119 parts (59%);
Technique 2 would have covered 142 parts (71%);
Technique 4 would have covered 156 parts (78%).

We investigated the corresponding percentages for the sample parts, using the experience of the first 12 months and predicting for the year from the 13th through the 24th month. In this case, the ratio of flying-hours in the prediction period to flying-hours in the experience period is a little more than 2. We estimate that:

Technique 1 would have covered 115 parts (42%);
Technique 2 would have covered 206 parts (76%);
Technique 4 would have covered 224 parts (82%).

There were 144 parts for which no demand was experienced in the first 12 months. Of these, 40 also showed no demand in the year which included months 13 through 24. There was an advantage in using 21 months' experience rather than 12, because there were fewer parts for which predictions had to be made with no demand experience. It is a difficult problem to estimate the future demand for a part for which there have been n successive months of zero demand. One suggestion we have is to assume a mean demand of one and a standard deviation of one, and use the probabilities of demand for 0, 1, 2, 3, ... as computed from the Poisson distribution.

REFERENCES

[1] See, for example, E. S. Pearson and H. O. Hartley (eds.), Biometrika Tables for Statisticians, Vol. 1, Cambridge University Press, 1954; Table 40, p. 203.

[2] Joseph Berkson, "Tables for Use in Estimating the Normal Distribution Function by Normit Analysis," Biometrika, Vol. 44, 1957, p. 411.

NOTE: The Appendix, consisting of Tables A1 through A6, will be found on the following 6 pages.

TABLE A-1

Average Monthly Error Of Predictions Using Six Techniques
(Engine Parts—02H1)

Exp. Months	Technique 1				Technique 2				Technique 3				Technique 4				Technique 5				Technique 6			
	1-12	1-15	1-18	1-21	1-12	1-15	1-18	1-21	1-12	1-15	1-18	1-21	1-12	1-15	1-18	1-21	1-12	1-15	1-18	1-21	1-12	1-15	1-18	1-21
Pred. Mos. / Part No.	13-24	16-27	19-30	22-33	13-24	16-27	19-30	22-33	13-24	16-27	19-30	22-33	13-24	16-27	19-30	22-33	13-24	16-27	19-30	22-33	13-24	16-27	19-30	22-33
103	8.3	4.8	3.7	2.5	10.4	6.1	4.7	3.2	13.9	9.1	7.4	5.6	12.2	7.2	5.6	3.9	225.	130.	83.	62.	95.	57.	41.	30.
104	0.5	-2.9	-2.8	-3.2	1.9	-1.9	-2.0	-2.6	3.3	-.4	-.4	-1.0	2.5	-1.4	-1.6	-2.2	50.	35.	23.	21.	9.9	7.8	5.3	4.8
108	0.1	-.4	0.4	-.1	1.4	0.4	0.4	0.5	2.6	1.6	1.7	2.0	2.0	0.8	0.7	0.9	3.0	2.0	6.9	13.3	-.9	-1.0	-.2	1.9
109	0.8	1.2	1.1	0.9	2.0	2.1	1.7	1.4	3.1	3.4	3.0	2.6	2.5	2.5	2.1	1.7	32.	26.	21.	19.	5.3	6.8	6.1	5.7
110	4.9	1.9	1.3	0.9	6.7	3.0	2.2	1.7	9.5	5.4	4.6	3.9	8.2	3.9	3.0	2.3	102.	57.	38.	35.	31.	22.	14.	12.
111	7.0	4.2	3.2	2.6	9.0	5.3	4.1	3.3	12.1	7.9	6.4	5.5	10.6	6.3	4.8	3.9	42.	44.	34.	35.	9.3	15.	12.	14.
112	-22.1	-21.7	-5.6	6.6	-20.2	-20.5	-3.8	8.3	-17.9	-17.6	0.9	13.6	-19.1	-19.4	-2.4	9.6	152.	90.	101.	215.	47.	25.	69.	168.
113	-.6	-.7	-1.2	-.7	-.1	-.4	-.9	-.4	0.5	0.1	-.4	0.3	0.1	-.3	-.8	-.3	--	--	-1.1	-.6	--	--	-1.3	-.9
149	-4.4	-3.4	-3.7	-5.6	-3.2	-2.3	-2.8	-4.8	-2.1	-.5	-.5	-2.5	-2.7	-1.7	-2.1	-4.2	0.7	4.3	8.9	12.	-5.8	-3.7	-1.9	-1.0

TABLE A-2

Average Monthly Error* Of Predictions Using Seven Techniques
(Airframe Components—1AFG)

Exp. Months	Technique 1				Technique 2				Technique 3				Technique 4				Technique 5				Technique 6				Technique 7			
	1-12	1-15	1-18	1-21	1-12	1-15	1-18	1-21	1-12	1-15	1-18	1-21	1-12	1-15	1-18	1-21	1-12	1-15	1-18	1-21	1-12	1-15	1-18	1-21	1-12	1-15	1-18	1-21
Pred. Mos / Part No.	13-24	16-27	19-30	22-33	13-24	16-27	19-30	22-33	13-24	16-27	19-30	22-33	13-24	16-27	19-30	22-33	13-24	16-27	19-30	22-33	13-24	16-27	19-30	22-33	13-24	16-27	19-30	22-33
201	-.5	-.04	0.02	0.1	0.2	0.6	0.6	0.5	0.6	1.3	1.3	1.3	0.5	0.8	0.7	0.7	--	0.6	0.4	0.3	--	-.3	-.2	-.1	--	--	--	--
202	-.1	-.2	0.1	0.6	-.04	0.4	0.7	1.1	0.5	1.4	1.9	2.3	0.3	0.8	1.1	1.4	--	3.8	3.9	3.8	--	0.7	1.1	1.4	--	--	--	--
203	-.6	-1.1	-1.3	-2.1	0.1	-.5	-.8	-1.7	0.5	.2	.1	-.8	0.3	-.3	-.6	-1.5	--	0.8	0.4	0.7	--	-1.0	-1.3	-1.5	--	--	--	--
204	3.2	3.5	2.6	1.7	5.1	4.5	3.4	2.3	7.3	7.0	5.7	4.3	6.1	5.5	4.1	2.8	9.3	7.0	5.4	4.9	6.9	5.1	3.6	3.0	--	--	--	--
205	-.5	0.1	0.2	0.2	0.1	0.6	0.7	0.5	--	1.2	1.5	1.3	0.2	0.8	0.9	0.7	--	2.0	2.7	2.9	--	0.3	0.6	0.8	--	--	--	--
206	-.6	0.5	0.3	0.3	--	1.2	0.8	0.7	-.1	2.1	1.7	1.6	0.2	1.5	1.0	1.0	--	1.8	0.9	1.2	--	.2	-.1	0.1	--	--	--	--
208	-2.3	1.9	1.7	1.9	-1.6	2.9	2.5	2.4	-1.2	4.8	4.3	4.2	-1.3	3.5	3.0	2.9	--	4.4	4.7	5.4	--	1.6	2.3	3.4	--	--	--	--
209	-.8	-.5	-.3	-.2	-.2	-.1	0.2	0.1	-.2	0.5	0.8	0.9	--	0.1	0.4	0.3	--	--	0.7	1.1	--	--	-.1	.1	--	--	--	--
211	-.8	-4.7	-4.6	-3.2	0.2	-4.1	-4.1	-2.6	0.9	-3.5	-3.3	-1.2	0.5	-3.9	-3.9	-2.2	--	--	-3.4	-1.8	--	--	-4.8	-3.2	--	--	--	--
214	-.5	-.8	-.8	-.2	0.1	-.5	-.6	0.1	--	-.5	-.6	0.8	0.2	-.4	-.5	0.3	--	--	--	4.5	--	--	--	0.5	--	--	--	--
215	-2.1	-4.6	-4.9	-.7	0.7	-2.7	-3.2	1.0	5.0	1.6	1.2	5.9	0.9	-1.1	-1.8	2.7	328.	192.	148.	123.	190.	123.	66.	58.	--	--	--	--
216	-.4	-.4	-.1	-1.0	0.5	0.1	0.4	-.6	1.0	0.8	1.2	0.4	0.8	0.4	0.6	-.4	--	2.7	2.2	1.7	--	0.2	0.2	-.5	--	--	--	--
217	1.0	0.6	0.4	0.4	2.0	1.2	0.9	0.7	3.0	2.1	1.7	1.4	2.5	1.5	1.1	0.9	7.2	6.1	5.5	4.5	3.6	3.1	2.7	2.1	--	--	--	--
218	0.6	0.1	-.1	-.7	1.6	0.7	0.4	-.4	2.5	1.5	1.1	0.4	2.0	1.0	0.6	-.2	1.7	2.5	3.2	2.3	0.2	0.2	0.5	-.1	--	--	--	--
219	0.2	-.4	-.6	-.5	1.5	0.4	0.1	0.1	2.8	1.7	1.4	1.5	2.1	0.8	0.4	0.4	2.6	2.9	2.2	1.9	-.3	0.1	0.1	--	--	--	--	--
220	-.9	-1.1	--	0.9	0.1	-.5	0.7	1.5	1.0	0.4	1.9	3.0	0.5	-.2	1.0	1.9	1.3	1.3	2.4	3.6	-.9	-.8	0.	1.1	--	--	--	--
221	0.5	--	0.2	0.6	1.6	0.7	0.8	1.1	2.7	1.7	1.9	2.3	2.1	1.0	1.1	1.4	4.7	3.9	3.4	3.9	1.3	1.1	1.0	1.6	--	--	--	--
222	-1.4	2.0	4.1	3.1	0.4	3.4	5.3	4.2	2.5	6.4	8.5	7.2	1.4	4.5	6.3	5.0	3.0	4.1	6.6	5.8	-.1	1.9	5.1	4.6	-1.1	3.0	5.1	3.8
223	-6.4	-.4	1.1	1.3	-5.4	0.8	2.3	2.4	-4.5	3.5	5.3	5.5	-5.0	1.8	3.2	3.2	-1.4	21.4	30.1	37.1	-6.5	6.5	14.1	21.2	--	--	--	--
224	-.5	-.7	-2.6	-2.7	0.7	0.2	-1.9	-2.0	1.9	1.7	-.4	-1.6	1.3	0.7	-1.5	-1.6	2.2	1.2	.04	.2	-.7	-.7	-1.9	-1.9	--	--	--	--
226	0.9	0.4	0.6	0.6	2.0	1.1	1.2	1.1	3.1	2.0	2.2	2.1	2.6	1.4	1.5	1.3	5.8	4.7	4.3	4.6	2.2	1.9	1.7	2.1	--	--	--	--
227	-.4	-.5	-.5	0.6	0.3	--	-.2	1.0	0.7	0.3	0.2	1.9	0.6	0.1	-.1	1.2	--	--	--	1.8	--	--	--	0.6	--	--	--	--
228	-.5	-.2	--	--	0.2	0.4	0.5	0.4	0.6	1.1	1.4	1.3	0.5	0.6	0.8	0.6	--	.9	1.5	1.4	--	-.2	0.2	0.2	--	--	--	--
229	-.3	0.2	0.1	--	0.2	0.7	0.6	0.4	0.2	1.4	1.2	1.0	0.4	1.0	0.8	0.5	--	--	10.9	10.7	--	--	1.9	2.0	--	--	--	--
232	--	0.2	-.1	-.1	0.8	0.8	0.4	0.3	1.4	1.6	1.2	1.1	1.1	1.1	0.6	0.5	--	1.2	1.4	1.3	--	0.	0.2	0.1	--	--	--	--
233	0.1	0.2	--	0.1	1.0	0.7	0.4	0.5	1.5	1.4	1.0	1.1	1.3	1.0	0.6	0.6	--	--	--	1.2	--	--	--	0.2	--	--	--	--
234	4.8	3.2	2.3	2.1	6.9	4.8	3.5	3.0	10.3	8.0	6.5	5.9	8.7	6.0	4.4	3.7	62.	64.	62.	49.	34.	45.	49.	37.	--	--	--	--
235	-.8	-.9	-.7	--	--	-.5	-.2	0.5	0.6	0.1	0.5	1.7	0.4	-.3	--	0.8	--	--	0.4	1.7	--	--	-.6	0.	--	--	--	--
240	0.7	0.9	0.9	1.0	1.9	1.7	1.6	1.6	3.2	3.1	2.9	3.0	2.6	2.2	2.0	1.9	6.4	10.8	9.6	10.1	0.1	2.0	2.2	2.6	--	--	--	--
242	-.6	0.8	0.6	0.5	--	1.5	1.2	0.9	-.1	2.5	2.1	1.8	0.2	1.8	1.4	1.1	--	4.0	1.8	1.0	--	1.0	0.5	0.3	--	--	--	--
243	-1.0	1.9	1.4	1.0	-.1	2.9	2.2	1.7	0.6	4.7	3.9	3.2	0.3	3.5	2.7	2.0	--	5.9	3.8	2.8	--	1.8	1.2	0.8	--	--	--	--
244	0.8	1.8	1.6	1.2	2.2	2.8	2.3	1.8	3.6	4.4	3.9	3.3	2.8	3.3	2.7	2.1	6.8	5.7	4.3	3.6	4.4	3.8	2.6	1.9	--	--	--	--
245	-.1	0.8	0.8	0.2	0.8	1.5	1.4	0.7	1.6	2.6	2.6	1.7	1.2	1.9	1.7	0.9	--	4.6	4.8	4.0	--	1.8	2.2	1.7	--	--	--	--
247	-.5	-.6	-.5	-.2	0.2	-.1	-.1	0.2	0.6	0.4	0.5	1.1	0.5	0.1	0.1	0.5	--	--	1.1	1.8	--	--	-.2	0.1	--	--	--	--
249	0.9	0.3	0.2	0.5	2.5	1.3	1.0	1.2	4.2	2.9	2.8	3.0	3.3	1.8	1.5	1.7	45	32	25	25	8.0	8.8	7.5	7.9	--	--	--	--

*For Techniques 1-4, the dashes indicate that Average Monthly Error was close to zero, for Techniques 5-7, the dash means that no prediction was made.

TABLE A-3

Average Monthly Error of Predictions Using Six Techniques
(Gunnery Components—FED. 1270)

Exp. Months / Pred. Mos. / Part No.	Technique 1				Technique 2				Technique 3				Technique 4				Technique 5				Technique 6			
	1-12	1-15	1-18	1-21	1-12	1-15	1-18	1-21	1-12	1-15	1-18	1-21	1-12	1-15	1-18	1-21	1-12	1-15	1-18	1-21	1-12	1-15	1-18	1-21
	13-24	16-27	19-30	22-33	13-24	16-27	19-30	22-33	13-24	16-27	19-30	22-33	13-24	16-27	19-30	22-33	13-24	16-27	19-30	22-33	13-24	16-27	19-30	22-33
301	-.8	-3.7	-5.4	-5.5	0.2	-3.1	-4.9	-5.1	0.9	-2.5	-4.0	-4.1	.5	-2.9	-4.6	-4.8	--	--	-2.5	-2.9	--	--	-4.7	-4.9
302	-.6	-.5	-.3	-.1	0.2	0.03	0.05	0.2	0.5	0.5	0.6	0.9	.4	0.2	0.2	0.4	--	--	--	1.5	--	--	--	0.2
304	-.6	-.6	-.6	-.2	-.1	-.3	-.4	0.06	0.5	0.2	0.2	0.8	.1	-.2	-.3	0.2	--	--	--	--	--	-.3	-.3	-.02
305	-.1	-.4	-.4	-.3	0.6	0.03	-.07	0.04	1.0	0.4	0.4	0.7	.8	0.2	0.1	0.2	--	--	--	1.4	--	--	--	-.08
310	-1.1	-1.9	-3.1	-2.4	-.3	-1.5	-2.8	-2.0	0.04	-1.1	-2.4	-1.1	-.09	-1.3	-2.6	-1.8	--	--	--	-.9	--	--	--	-2.5
313	-.1	0.3	-.2	-.5	0.8	1.0	0.4	-.1	1.6	2.0	1.3	0.7	1.2	1.3	0.6	0.05	--	1.6	0.8	0.4	--	.2	-.3	-.5
314	-2.4	-3.1	-4.6	-5.6	-1.5	-2.4	-4.0	-5.0	-1.0	-1.4	-3.0	-3.6	-1.2	-2.1	-3.7	-4.7	--	0.4	-.2	-1.4	--	-2.4	-2.9	-3.7
317	-1.7	-1.6	-1.0	-1.2	-.8	-.8	-.3	-.6	-.02	0.5	1.2	1.0	-.4	-.3	0.2	-.2	-.1	0.9	2.3	2.0	-1.7	-1.3	.08	0.2
318	-.5	-.7	-.8	-1.6	0.1	-.1	-.4	-1.2	0.02	0.5	0.5	-.5	0.2	0.1	-.1	-1.1	--	--	1.7	1.5	--	--	-.6	-1.0
320	-2.1	-3.9	-6.0	-6.3	-1.5	-3.5	-5.6	-5.7	-1.6	-3.1	-5.0	-4.5	-1.3	-3.3	-5.4	-5.4	--	--	--	-3.8	--	--	--	-6.1
321	-1.8	-2.2	-2.7	-2.2	-1.2	-1.9	-2.4	-1.6	-1.2	-1.9	-1.9	-.5	-1.0	-1.8	-2.2	-1.4	--	--	--	-.2	--	--	--	-2.2
322	-5.3	-11.7	-14.5	-13.5	-4.8	-11.3	-14.1	-12.8	-4.8	-11.4	-13.5	-10.8	-4.6	-11.2	-13.9	-12.2	--	--	--	-11.8	--	--	--	-13.9
323	-2.5	-3.5	-5.1	-5.3	-1.9	-3.0	-4.5	-4.7	-2.0	-2.3	-3.5	-3.3	-1.8	-2.7	-4.2	-4.4	--	--	--	-2.0	--	--	--	-4.3
325	-1.0	-.4	0.02	0.3	-.4	0.2	0.4	0.7	0.8	1.5	1.8	2.0	-.03	0.5	0.7	0.9	--	--	--	--	-1.4	-.4	-.01	0.3
326	-1.2	-1.1	-.1	-.2	-.7	-.6	0.4	0.3	-.7	-.3	1.3	1.3	-.5	-.5	0.7	0.5	--	--	.04	0.6	--	--	-.3	-.04
327	-2.0	-3.2	-3.7	-2.1	-1.4	-2.9	-3.3	-1.5	-1.5	-2.9	-2.8	-.1	-1.3	-2.8	-3.1	1.1	--	--	--	-.00	--	--	--	-2.0
329	-1.9	0.5	0.6	0.8	4.8	2.6	2.2	2.0	4.2	4.0	3.9	3.7	6.7	3.6	3.0	2.6	--	1.4	2.3	1.3	--	-1.5	-1.0	-.3
331	-2.2	-3.7	-5.0	-4.9	-1.7	-3.3	-4.7	-4.6	-1.7	-3.4	-4.8	-4.0	-1.5	-3.2	-4.7	-4.4	--	--	--	--	--	--	--	-5.0
334	-4.2	-6.4	-7.8	-7.0	-3.7	-6.0	-7.2	-6.3	-3.7	-5.6	-6.1	-4.5	-3.6	-5.8	-6.9	-5.8	--	-5.0	-5.9	-4.6	--	-6.2	-7.8	-6.7
335	-1.2	-2.5	-3.1	-3.5	-.6	-2.2	-2.8	-3.2	-.7	-2.2	-2.9	-2.6	-.4	-2.1	-2.8	-3.1	--	--	--	-2.5	--	--	--	-3.5
338	-2.1	-3.2	-5.2	-5.2	-1.5	-2.8	-4.7	-4.7	-1.6	-2.9	-4.2	-3.6	-1.3	-2.7	-4.6	-4.5	--	--	--	-2.9	--	--	--	-4.9
339	-2.8	-2.4	-2.0	-1.8	-2.3	-1.8	-1.5	-1.2	-.3	-1.0	-.4	0.2	-2.1	-1.5	-1.2	-.8	--	--	--	0.8	--	--	--	-1.8
340	-.8	-1.2	-1.4	-.8	-.2	-.9	-1.2	-.2	-.2	-.9	-1.2	0.8	0.01	-.8	-1.1	-.02	--	--	--	-.2	--	--	--	-1.0
341	-1.2	-1.3	0.1	0.01	-.6	-.9	0.7	0.5	-.6	-.5	1.9	1.6	-.4	-.7	1.1	0.8	--	--	2.5	4.1	--	--	0.1	1.1
343	-4.2	-6.9	-8.1	-8.1	-3.7	-6.4	-7.5	-7.3	-3.7	-5.8	-6.3	-5.6	-3.5	-6.2	-7.2	-6.9	--	--	-4.0	-1.0	--	--	-7.6	-5.9
344	0.2	0.2	0.4	0.3	1.2	.8	1.0	0.7	1.9	1.6	1.8	1.5	1.5	1.1	1.2	0.9	--	3.0	3.3	4.0	--	0.2	0.5	0.6
345	0.4	0.4	0.3	0.3	1.3	1.0	0.7	0.6	2.1	1.7	1.4	1.4	1.7	1.2	0.9	0.8	4.0	3.6	3.0	2.7	1.0	0.9	0.8	0.7
348	-.2	0.2	0.1	0.3	0.6	0.9	0.6	0.7	1.2	1.7	1.4	1.6	1.0	1.2	0.8	1.0	--	4.8	4.5	4.3	--	1.0	0.9	1.0
349	-.2	0.01	-.2	-.01	0.5	0.6	0.2	0.3	0.9	1.2	0.8	1.0	0.8	0.8	0.4	0.5	--	--	--	4.3	--	--	--	1.0
350	1.0	0.5	0.6	0.7	2.2	1.3	1.3	1.3	3.5	2.5	2.4	2.4	2.8	1.7	1.6	1.5	2.44	3.47	4.88	4.73	0.1	0.5	2.0	2.1

TABLE A-4

Average Monthly Error of Predictions Using Seven Techniques
(Bombing Fire Control Components—1280)

Exp. Months / Pred. Mos / Part No.	Technique 1				Technique 2				Technique 3				Technique 4				Technique 5				Technique 6				Technique 7			
	1-12	1-15	1-18	1-21	1-12	1-15	1-18	1-21	1-12	1-15	1-18	1-21	1-21	1-15	1-18	1-21	1-12	1-15	1-18	1-21	1-12	1-15	1-18	1-21	1-12	1-15	1-18	1-21
	13-24	16-27	19-30	22-33	13-24	16-27	19-30	22-33	13-24	16-27	19-30	22-33	13-24	16-27	19-30	22-33	13-24	16-27	19-30	22-33	13-24	16-27	19-30	22-33	13-24	16-27	19-30	22-33
402	-8.2	-11.7	-12.5	-11.1	-6.6	-10.5	-11.3	-9.9	-4.8	-7.8	-8.2	-6.2	-5.7	-9.5	-10.4	-8.9	-6.7	-10.0	-11.0	-11.2	-8.6	-12.8	-13.1	-12.4	--	--	--	--
403	0.1	0.6	0.4	0.6	1.1	1.3	1.0	1.1	2.0	2.3	2.0	2.1	1.5	1.6	1.2	1.3	3.1	3.9	3.3	2.7	0.5	1.4	1.2	1.0	--	--	--	--
405	-2.0	-2.0	-3.0	-2.7	-.9	-1.1	-2.2	-2.1	0.2	0.4	-.6	-.4	-.4	-.6	-1.8	-1.7	0.	1.5	1.2	1.2	-2.5	-1.7	-1.6	-1.1	--	--	--	--
408	-3.8	-3.0	-3.1	-1.2	-2.8	-2.0	-2.3	-.6	-2.0	-.4	-.7	1.6	-2.4	-1.5	-1.9	0.	5.0	8.7	9.9	12.4	-3.0	-.1	1.3	4.6	--	--	--	--
416	-1.7	-2.0	-.8	-.1	-1.1	-1.5	-.2	0.4	-1.2	-1.0	0.6	1.6	-.9	-1.3	0.1	0.6	--	--	1.9	3.6	--	--	-.6	0.9	--	--	--	--
417	-4.6	-5.3	-5.0	-2.5	-4.0	-4.8	-4.5	-1.8	-4.1	-4.3	-3.6	-.1	-3.9	-4.6	-4.3	-1.5	--	--	-3.9	-.3	--	--	-5.0	-2.2	--	--	--	--
418	-8.5	-3.1	.9	1.3	-7.0	-1.7	2.3	2.3	-5.3	1.3	5.8	5.8	-6.2	-.6	3.3	3.2	-3.0	-.7	2.0	2.0	-6.1	-2.8	.6	.8	-7.6	-2.8	-.2	0.6
421	-4.7	-4.6	-6.5	-2.0	-3.3	-3.6	-5.7	-1.2	-1.8	-1.9	-4.0	1.6	-2.6	-3.1	-5.3	-.5	-4.3	-1.9	-4.2	-2.0	-5.8	-4.5	-6.4	-4.0	-.8	-.6	-1.6	-1.2
422	-.9	-1.2	-1.5	-1.1	0.4	-.3	-.8	-.5	1.8	1.2	0.8	1.5	1.1	0.2	-.4	0.06	5.0	3.3	2.9	3.2	-.2	-.5	-.9	-.8	--	--	--	--
423	1.2	0.8	0.6	0.4	2.4	1.5	1.1	0.8	3.5	2.6	2.1	1.6	2.9	1.9	1.4	1.0	4.1	1.8	0.7	0.2	1.8	0.8	0.4	0.1	--	--	--	--
424	-1.2	-1.5	-1.5	0.5	0.2	-.6	-.8	1.1	1.6	0.9	0.9	3.3	0.8	-.1	-.3	1.7	8.3	2.8	2.4	4.0	1.0	-.6	-.6	0.8	--	--	--	--
425	0.6	0.3	0.3	0.2	1.6	0.8	0.8	0.5	2.5	1.6	1.5	1.2	2.0	1.1	1.0	0.7	4.2	3.6	2.9	2.5	1.2	1.2	0.9	0.6	--	--	--	--
427	0.7	0.2	0.5	0.2	2.0	1.0	1.2	0.7	3.4	2.3	2.6	2.1	2.6	1.4	1.5	1.0	6.5	5.1	4.6	3.7	3.1	2.3	2.1	1.4	--	--	--	--
429	-.04	-.3	-.2	-.1	0.8	0.3	0.3	0.2	1.4	0.9	1.0	1.0	1.1	0.5	0.5	0.4	--	--	5.8	5.8	--	--	0.7	0.7	--	--	--	--
430	-.8	-1.1	-.7	-.6	-.1	-.7	-.2	-.2	0.3	-.4	0.6	0.8	0.2	-.6	0.02	0.05	--	--	0.3	2.2	--	--	-.8	0.1	--	--	--	--
431	-2.3	-1.6	-1.8	-.4	-.9	-.6	-.9	0.4	0.5	1.1	0.9	2.7	-.3	-.1	-.5	1.0	2.8	2.9	2.5	2.4	-.9	0.2	0.5	0.7	--	--	--	--
433	-.1	-.2	0.	0.6	1.6	0.8	0.8	1.4	3.7	3.2	3.3	3.8	2.6	1.7	1.7	2.0	5.4	4.0	3.9	3.6	2.7	2.0	2.3	2.1	0.3	1.1	0.5	1.0
436	-.4	-.9	-.4	-1.4	0.7	-.1	0.3	-.9	1.8	1.1	1.7	0.4	1.2	0.3	0.6	-.6	3.2	2.9	3.2	2.6	0.1	-.1	0.4	0.2	--	--	--	--
440	0.4	0.1	0.7	0.2	1.6	0.8	1.3	0.7	2.8	1.8	2.5	1.8	2.2	1.1	1.6	0.9	4.5	3.8	3.4	2.4	1.3	1.4	1.4	0.8	--	--	--	--
441	1.7	0.3	0.3	1.7	3.5	1.4	1.1	2.4	5.8	3.5	3.2	4.6	4.6	2.1	1.7	2.9	17.0	15.4	15.1	15.0	7.6	8.6	8.8	10.1	--	--	--	--
442	-5.2	-4.2	-4.3	0.5	-4.3	-3.4	-3.5	1.2	-3.4	-2.0	-1.9	3.8	-3.8	-2.9	-3.1	1.9	0.2	0.3	1.8	7.7	-4.3	-3.7	-3.3	1.2	--	--	--	--
443	-15.2	-12.9	-13.4	5.5	-12.8	-10.9	-11.7	7.3	-8.8	-6.6	-7.2	13.1	-10.7	-9.4	-10.4	8.6	-14.2	-13.8	-15.6	-4.6	-15.6	-13.8	-16.1	-5.5	-2.4	-3.4	-3.2	-3.7
447	0.8	0.2	0.9	-.1	2.2	1.0	1.7	0.5	3.7	2.4	3.3	2.0	2.9	1.5	2.1	0.9	3.7	4.1	4.4	3.7	0.5	1.4	1.9	1.4	--	--	--	--
450	-3.4	-2.8	-1.7	1.8	-2.4	-2.0	-.9	2.5	-1.6	-.9	0.7	4.8	-2.0	-1.6	-.5	3.1	-1.1	-1.3	0.2	3.5	-3.2	-2.8	-2.0	1.1	--	--	--	--

TABLE A-5

Average Monthly Error of Predictions Using Seven Techniques
(Communications Equipment—Federal Class 5800, AF 16A, $16K_1$, $16K_2$)

Exp. Mos.	Technique 1				Technique 2				Technique 3				Technique 4				Technique 5				Technique 6				Technique 7			
	1-12	1-15	1-18	1-21	1-12	1-15	1-18	1-21	1-12	1-15	1-18	1-21	1-12	1-15	1-18	1-21	1-12	1-15	1-18	1-21	1-12	1-15	1-18	1-21	1-12	1-15	1-18	1-21
Pred. Mos. / Part No.	13-24	16-27	19-30	22-33	13-24	16-27	19-30	22-33	13-24	16-27	19-30	22-33	13-24	16-27	19-30	22-33	13-24	16-27	19-30	22-33	13-24	16-27	19-30	22-33	13-24	16-27	19-30	22-33
503	-1.1	0.1	0.4	0.5	-.4	0.8	1.1	1.1	-.05	1.9	2.4	2.4	-.2	1.2	1.4	1.4	--	2.8	4.4	4.7	--	0.2	1.6	2.5	--	--	--	--
506	0.4	-.3	1.5	2.0	2.4	0.9	2.6	2.9	4.8	3.5	5.4	5.7	3.5	1.9	3.5	3.7	4.6	2.8	4.4	4.3	2.3	0.6	2.6	2.6	0.2	0.2	1.9	2.8
507	-.9	-.1	3.8	4.4	1.1	1.2	5.1	5.5	3.5	3.9	8.3	8.7	2.2	2.2	6.1	6.4	1.8	2.3	5.0	4.9	-1.8	-.2	3.2	3.0	--	--	--	--
508	10.5	4.2	0.1	-2.0	13.4	6.1	1.6	-.7	17.9	10.1	5.4	3.1	15.7	7.5	2.8	0.3	1.7	0.4	-2.2	-4.2	-.5	-1.3	-3.5	-5.7	3.2	1.3	-1.8	2.2
513	-1.8	-2.4	-2.1	-2.4	-.6	-1.3	-1.2	-1.6	0.5	0.5	1.2	0.7	-.1	-.7	-.4	-1.0	0.7	0.6	1.2	1.2	-2.2	-2.2	-.9	-.4	--	--	--	--
515	-2.2	-2.8	1.1	0.7	-1.1	-1.9	2.1	1.5	0.1	-.3	4.7	4.0	-.5	-1.4	3.0	2.2	1.1	0.06	3.2	3.3	-1.7	-2.4	0.8	1.3	--	--	--	--
516	-.2	-.02	-1.3	0.2	1.6	1.1	-.2	1.1	3.7	3.7	2.6	4.0	2.6	2.1	.7	1.9	4.0	3.1	2.3	2.6	0.5	2.0	1.0	1.0	-.7	0.1	-1.8	-.4
518	2.8	0.03	-.1	-2.1	4.7	1.1	0.9	-1.3	0.3	-.03	-.2	0.5	5.9	2.1	1.8	-.6	4.2	2.4	1.8	0.2	0.7	-.5	-.7	-1.6	--	--	--	--
519	-1.6	-.6	-.6	0.1	-.1	0.5	0.5	1.0	1.6	3.1	3.3	3.8	0.7	1.5	1.4	1.8	3.5	2.6	2.1	2.4	-.3	0.3	0.3	0.9	-1.4	-.5	-1.1	0.2
523	1.1	0.6	0.6	0.5	2.2	1.3	1.2	0.9	3.3	2.2	2.1	1.8	2.7	1.6	1.4	1.1	--	--	5.2	4.4	--	--	2.5	2.0	--	--	--	--
524	-.5	0.01	-1.3	-3.5	0.8	1.0	-.4	-2.8	2.1	2.7	1.5	-1.0	1.4	1.6	0.2	-2.3	5.2	4.7	3.2	0.8	1.4	1.9	1.4	-1.0	--	--	--	--
528	-1.0	-1.2	-.9	-.6	-.3	-.6	-.3	-1.1	0.1	0.1	0.8	1.1	--	-.4	0.01	0.2	--	--	0.7	1.5	--	--	-.8	-.2	--	--	--	--
544	0.05	0.5	0.2	0.1	0.9	1.1	0.7	0.5	1.5	2.0	1.5	1.3	1.2	1.4	0.9	0.7	--	1.6	1.6	1.5	--	0.3	0.3	0.2	--	--	--	--
546	-.06	-.6	0.8	0.5	0.9	-.04	1.5	1.0	1.8	0.7	2.8	2.2	1.4	0.2	1.8	1.3	1.2	0.6	2.6	4.1	-.5	-.7	0.4	1.3	--	--	--	--
548	4.7	2.7	7.5	5.2	7.5	4.6	9.2	6.7	11.9	8.8	13.7	11.1	9.8	6.2	10.6	7.9	1.3	0.3	4.0	1.9	-1.1	-1.8	2.4	0.8	1.8	1.6	4.6	5.1
550	-1.4	-.6	0.1	0.8	-.3	0.3	0.9	1.4	0.8	1.8	2.6	3.5	0.2	0.8	1.4	2.0	3.8	0.9	0.4	1.6	0.8	-.6	-.3	0.6	--	--	--	--

TABLE A-6

Average Monthly Error of Predictions Using Seven Techniques
(Accessories, Equipment and Parts—03B, C, F, I)

Exp. Months	Technique 1				Technique 2				Technique 3				Technique 4				Technique 5				Technique 6				Technique 7			
	1-12	1-15	1-18	1-21	1-12	1-15	1-18	1-21	1-12	1-15	1-18	1-21	1-12	1-15	1-18	1-21	1-12	1-15	1-18	1-21	1-12	1-15	1-18	1-21	1-12	1-15	1-18	1-21
Pred. Mos. / Part No.	13-24	16-27	19-30	22-33	13-24	16-27	19-30	22-33	13-24	16-27	19-30	22-33	13-24	16-27	19-30	22-33	13-24	16-27	19-30	22-33	13-24	16-27	19-30	22-33	13-24	16-27	19-30	22-33
601	-1.2	-1.5	-2.3	-5.4	5.5	-.5	-1.0	-4.4	4.9	-.6	0.3	-3.2	7.4	-.2	-.4	-4.0	--	--	0.8	-.9	--	--	-3.6	-5.7	--	--	--	--
603	-.05	-.2	-.07	-.02	0.4	0.2	0.2	0.2	1.0	0.8	0.9	0.8	0.6	0.3	0.3	0.3	--	2.8	2.2	1.8	--	0.2	0.2	0.3	--	--	--	--
606	-.4	-3.5	-6.9	-5.4	1.1	-2.6	-6.2	-4.7	2.6	-1.3	-4.8	-2.5	1.8	-2.2	-5.8	-4.1	48.4	14.7	2.0	5.7	10.4	1.4	-4.6	-2.4	--	--	--	--
608	-2.2	-3.8	-6.0	-6.8	-1.7	-3.4	-5.5	-6.3	-1.7	-3.0	-4.7	-4.9	-1.5	-3.2	-5.3	-6.0	--	--	6.0	7.0	--	--	-3.0	-2.0	--	--	--	--
611	6.5	8.5	9.8	8.7	9.7	10.9	11.9	10.5	14.7	16.1	17.2	15.6	12.3	12.8	13.5	11.8	17.4	27.7	32.2	29.0	-.01	16.4	23.7	22.3	--	--	--	--
612	0.6	0.7	0.5	0.4	1.6	1.3	1.0	0.7	2.5	2.2	1.8	1.4	2.0	1.6	1.2	0.9	7.5	4.4	3.1	2.0	1.4	1.3	0.9	0.6	--	--	--	--
615	0.4	0.2	0.1	0.05	0.9	0.5	0.3	0.2	1.6	1.1	0.9	0.7	1.1	0.7	0.4	0.3	1.3	0.8	0.5	0.5	0.1	-.00	-.03	-.03	--	--	--	--
616	-.3	-.5	0.6	0.6	0.7	0.02	1.1	1.0	1.4	0.7	2.2	2.1	1.0	0.2	1.4	1.2	--	--	6.2	7.8	--	--	1.5	2.6	--	--	--	--
617	-.5	-.9	-.6	-.6	0.1	-.5	-.2	-.3	0.02	-.1	0.5	0.4	0.2	-.3	-.00	-.1	--	--	0.7	0.8	--	--	-.4	-.4	--	--	--	--
618	-.4	-.1	-.6	-.8	0.2	0.5	-.2	-.4	0.1	1.1	0.5	0.3	0.3	0.7	-.00	0.3	--	--	5.0	4.3	--	--	-.3	0.05	--	--	--	--
619	-2.3	-5.6	-7.0	-7.5	-1.8	-5.2	-6.7	-7.1	-1.8	-5.3	-6.8	-6.1	-1.6	-5.1	-6.7	-6.8	--	--	--	-5.6	--	--	--	-7.6	--	--	--	--
620	5.6	5.7	5.8	3.6	7.6	7.1	7.0	4.6	11.0	10.3	10.2	7.5	9.4	8.3	8.0	5.4	32.8	26.0	25.0	21.6	24.8	20.7	17.8	15.1	--	--	--	--
621	-.7	-.9	-.7	-.6	-.1	-.6	-.4	-.2	-.2	-.6	0.1	0.5	0.1	-.5	-.2	-.02	--	--	--	0.9	--	--	--	-.6	--	--	--	--
623	-3.7	-5.3	-3.4	-2.6	-3.1	-4.8	-2.7	-1.9	-3.1	-4.1	-1.2	-.2	-2.9	-4.6	-2.3	-1.5	--	--	-.2	1.1	--	--	-3.0	-1.1	--	--	--	--
625	0.8	0.6	0.3	-.1	1.3	1.0	0.7	0.3	2.6	2.2	1.8	1.4	1.6	1.3	0.9	0.5	0.6	0.4	0.3	-.1	0.4	0.3	0.2	-.2	1.0	0.9	0.6	-.3
626	-.6	-.6	-.9	-.02	-.01	-.2	-.6	0.4	-.06	-.3	-.7	1.3	0.2	-.1	-.6	0.6	--	--	--	-.1	--	--	--	-.2	--	--	--	--
627	-5.5	-.6	-.3	0.4	-4.6	0.5	0.6	1.2	-3.9	3.0	3.2	3.7	-4.2	1.4	1.5	1.8	--	-.9	1.2	3.0	--	-2.4	-.4	1.8	--	--	--	--
628	-8.1	-10.1	-12.8	-10.7	-7.5	-9.5	-12.1	-9.8	-7.6	-8.6	-10.7	-7.1	-7.3	-9.2	-11.7	-9.1	--	-8.8	-11.5	-5.3	--	-10.2	-13.2	-8.8	--	--	--	--
629	-.3	-7.0	-5.7	-5.0	2.6	-5.1	-4.0	-3.5	7.2	-.8	0.5	1.1	5.0	-3.5	-2.6	-2.4	81.0	61.4	59.0	98.8	22.5	22.9	29.2	63.3	--	--	--	--
631	-5.4	-2.5	-1.3	-.5	-3.4	-.8	0.3	0.9	-.1	3.1	4.4	5.1	-1.7	0.6	1.6	2.1	57.0	77.5	94.3	108.3	0.2	23.7	43.2	59.0	--	--	--	--
633	-7.6	-8.2	-8.8	-6.8	-7.0	-7.6	-8.1	-6.0	-7.1	-6.6	-6.8	-3.7	-6.8	-7.2	-7.8	-5.4	-7.6	-7.1	-7.1	11.3	-7.6	-8.5	-9.0	-2.5	--	--	--	--
635	0.1	0.1	-.3	-.1	1.3	1.0	0.4	0.4	2.5	2.3	1.6	1.7	1.8	1.4	0.7	0.7	3.7	4.4	4.2	3.9	0.4	1.6	1.6	1.5	--	--	--	--
636	-.04	-.7	-.7	-1.3	1.0	0.00	-.02	-.7	2.0	1.0	1.3	0.5	1.5	0.3	0.3	-.4	16.8	11.4	14.8	19.4	1.0	0.7	1.8	8.3	--	--	--	--
637	3.4	-.6	-.6	-2.1	5.8	1.0	0.9	-.8	9.7	4.5	4.7	3.0	7.8	2.3	2.1	0.2	187.	79.	52.	59.	86.9	46.1	31.4	32.8	--	--	--	--
640	0.2	1.7	1.2	1.2	1.5	2.7	2.0	1.8	2.9	4.5	3.7	3.4	2.2	3.3	2.5	2.1	--	--	--	12.7	--	--	--	5.9	--	--	--	--
641	-.7	-.8	0.5	0.4	-.1	-.4	1.0	0.8	-.1	-.4	1.9	1.8	0.1	-.3	1.3	1.1	--	--	1.1	1.6	--	--	0.05	0.1	--	--	--	--
642	-1.4	-1.2	3.3	2.8	-.1	-.4	4.2	3.6	1.1	1.0	6.6	5.8	0.5	0.1	5.0	4.2	3.9	3.6	5.6	5.7	0.3	1.1	3.1	3.9	--	--	--	--
643	3.9	1.8	0.5	-.7	5.8	3.0	1.5	0.1	8.2	5.1	3.4	1.9	7.0	3.7	2.0	0.5	5.7	5.6	4.7	3.1	2.8	3.3	2.4	1.1	--	--	--	--
645	7.0	4.3	3.3	2.4	8.9	5.5	4.2	3.2	12.1	8.2	6.6	5.4	12.1	6.5	5.0	3.8	6.2	6.8	7.1	6.2	3.5	5.0	5.7	5.1	--	--	--	--
646	11.6	7.3	6.9	6.0	17.9	9.5	8.7	7.5	19.7	14.3	13.4	12.1	17.3	11.2	10.2	8.7	2.1	0.9	1.3	0.9	0.6	-.3	-.2	-.4	1.8	0.8	3.2	3.6
650	-.02	-1.0	0.2	0.5	1.4	0.00	1.1	1.2	3.0	1.6	3.1	3.3	2.2	0.5	1.7	1.8	24.5	14.9	14.7	16.8	8.6	5.6	6.5	8.5	--	--	--	--

REALISM IN EMPIRICAL LOGISTICS RESEARCH *

J. E. Hamilton

Logistics Research Project

This discussion, although oriented toward research in logistics, including the initial gathering of pertinent data and subsequent search for structure, is limited to elemental aspects of problems in simulation of functions. The two specific problems outlined - with suggestions for method of attack - are pertinent to Navy ships only: What load should be carried when the ship is deployable and available for a mission, and what provisions are necessary for back-up support? The load and support problems are, in turn, divisible into three subordinate considerations: manpower, facilities, and materials. The latter receives the most attention in this paper.

Because the present classification of materials is unsatisfactory, a new classification is suggested, based on three criteria: The nature of utilization, the status (as repairable or disposable), and the rate of utility exhaustion. The first two are important for classification; the third is the core of an empirical study.

In determining what is to be loaded into a ship, certain predictable items are considered: population, time, and usage rate. The Navy currently has no procedure for collecting good experience data on usage, although only specific experience is reducible to clerical statistics. A job order system is therefore proposed.

INTRODUCTION

I can devise no better introduction to this paper than a quote from Professor J. A. Huston:

> In thinking about war and its principles, theories, and applications, how do we arrive at these things? How do we arrive at the principles to guide our thinking on war? It seems to me mainly that it is through a study of history—noting the points of similarity and contrast to all conclusions, testing these conclusions, revising our conclusions in the light of any new evidence. These conclusions always must be tentative; the principles must always be open-minded. Even the best established principles should remain open to question. ..... we want to search for all kinds of examples in order to revise the principles where necessary, and reaffirm them if that is what the evidence points to. Experience is the raw material of imagination, and history is vicarious experience.
>
> We want to look at these things in perspective. In doing so, I think we should apply what you might call two of the fundamental laws of history, if it can be said to have any laws. These fundamental laws which are present in governing any situation are continuity and change. If someone says this situation is completely different from anything

* Research was supported by the Office of Naval Research.

> before, he cannot be right. When he says that this thing is exactly like a previous situation, that could not be right either. These two things, although contradictory, are going together—change and continuity.
>
> In this way we look, then, at times such as these when we are in the midst of profound change. Sometimes I think we get our thinking a little bit off center. Some people attempt to avoid the responsibilities of the present by retreating into the past in the name of tradition; others evade the demands of transition by trying to escape into some unreal world of the present without any past or an even more unreal world of the future with neither a present nor a past. But please remember the past is the present—it is our thinking now on the past. It is the experience we have gained in the past with the experience which we have now based on the past. It is our traditions, our attitudes, our prejudices now that have developed out of the past which govern our thinking.

I have neither the intention nor the competence to continue this paper at the elevated level at which it has started by borrowing from Professor Huston. I shall try perhaps to expand some of his thoughts in my own words and in the context of my own thoughts.

It must be clearly understood that, from now on, anything that appears in this paper pertains, as far as the author is concerned, only to Naval ships. Most of this paper will be critical, constructively it is hoped, of what the author has come to believe is general thinking. Any such generalization is, of course, faulty and untenable, just as many of those which follow may be. I have neither the knowledge nor the experience to justify these beyond the scope of those who think and act in the field of Naval ships.

The above statement was made because I feel that too frequently an effort is made to extend the applicability of some fact which concerns one service only to all services and to all conditions and even to all circumstances. I feel specifically that many in the Department of Defense have done this in their official dicta, to the general detriment of defense administration. Each of the three services is so unique that one of the primary purposes of the Department of Defense should be to recognize singularities and to decentralize, to the military department level, all functions which fall in the unique areas.

The above is my impression. It is based on feelings which are not too well documented because I have not undertaken the research necessary to do this. The one example which I would mention, and which probably will be generally contested, is the Federal stock-number system. The published cost of doing this job when it was reported as completed was $150,000,000. This may represent a third of the total cost, direct and indirect. How possible savings even approaching this figure can ever be realized is impossible to visualize. The technical value of the system which would permit easy cross-servicing by clerks, without the aid of technicians, may apply to as many as 10,000 items out of the hundreds of thousands affected. By blindly applying the system to everything and then limiting its application so that it doesn't apply to everything, the Navy is saddled with a cumbersome system from which it cannot possibly derive any comparable advantage. Furthermore, the job was not finished. It will never be finished. It is a continuing job which has created additional permanent jobs which are permanent cost additions. A reliable and responsible person has told me that the Navy had changed 300,000 stock numbers in the first eight months of 1959.

Not all of this unbelievably large number of changes can be charged to the Federal stock numbering system; but many can be, either directly or as a result of the unsettled situation brought about by the system.

In general, the material for this paper was derived from the author's experience over an eight-year period in his rather intimate connection with the Data Collection Programs of the Logistics Research Project. During this time myriads of data in the form of supply documents, repair requests, and job orders were reviewed, a large volume of correspondence was carried out, and numerous visits to ships and supporting shore activities were made.

The general criticism which I will make is that formal procedures which govern the Navy's stocking of its supply system and the distribution of its stocks are dominated by considerations of the needs of the fixed elements (shore establishments) in the Supply System, but the actual operation of ship supply and resupply is dominated by the people who man the ships. There appears to be no adequate organizational meeting place between the two. This goad is inserted early in the paper to arouse antagonisms. As the paper develops, any misunderstanding which the simple statement of the criticism may engender will be either converted to understanding or clarified by specifics.

The Naval ship is a completely self-contained military unit, but the adequacy of the containment is limited by time and capabilities. The time limitation is dominant in the minds of most logistics planners and operators. However, some of the planners do consider the limitations of manpower (both quantitatively and qualitatively) and facilities. There is a second-level limit in each of these which is imposed by a different kind of time limitation.

The simple time-limitation and experience factors permit fairly valid conclusions, adequate for planning, so that:

(1) If reasonably correct assumptions are made as regards the specifics of future steaming conditions (speeds and distance or hours underway), a ship can be independent of fuel resupply for a certain number of days to fit any policy of low fuel limits (normal and emergency).

(2) If reasonably correct assumptions are made as to rate of expending ammunition, as was feasible after the Korean operation settled down, a ship can be independent of ammunition resupply for a certain number of days. Since in many kinds of wars such predictive assumptions are impossible, because opportunity is a strong factor in the time and rate of shooting, ammunition supply and resupply present a unique problem which will not be included in this paper.

(3) Just on the basis of experience, with no need for imaginative assumptions, a ship can be independent of food resupply on a normal basis for so many days and on an austere basis for so many more.

Each of these three material-categories—fuel, ammunition, and provisions—constitutes a problem in itself. They are thoroughly and generally understood throughout the Navy and elsewhere and will not be considered in detail in this paper.

Articles for resale—ship's store and clothing and small stores items—can be classed as morale-building, elements along with motion pictures, library books, and mail. Because of the close personal interest in these categories, they are seldom really neglected. Problems exist and new ones arise which require study, but most of these are unique to the category. Therefore, they also will not be considered within the scope of this paper.

In all other categories, demands which arise within a ship at any time can be considered in terms of three criteria: manpower, facilities, and commodities. In meeting a demand aboard ship, the supply officer must consider, consistently, whether one or more of these elements and time are available, in view of limitations inherent in the situation on his ship or because of time limits imposed by others. At any rate, the interdependence of the criteria will be a matter of human judgment. Decisions reached in each case will vary with personalities, even under identical circumstances.

One of the first things which came to our attention during the early stages of data collection was that the data which we were receiving from different sources had peculiarities of some kind which made every data source unique. Often these peculiarities became apparent to us when a Commanding Officer, a Supply Officer, or even a Supply Storekeeper had been changed. Thus, we noted variations among similar kinds of ships and from time to time, within the same ship. The author interprets this situation as the effect of personality, and he wishes to go on record as believing that this is the greatest single asset which the Navy possesses.

As it happens, the Navy's way of doing business recognizes the value of personality and by custom and practice protects it. This is not recognized officially in the great number of bureaucratic instructions and directives which are continually flowing from the administrative hierarchy. One could read and study these documents and could form a clear mental picture of exactly what was being done in all ships and stations. However, when he goes into the field with this picture in mind, he encounters differences—some subtle, some obvious. The instruction and the directive and the regulation, as they are policed, do not stultify imagination, initiative, energy, mission-consciousness, nor any of the other good human characteristics which are mainly responsible for the over-all effectiveness of the Navy. Much as a few true bureaucrats would like to see it done, the Navy cannot be regimented to the degree that reduces individuals to robots and completely standardizes their thinking and action.

The problem which underlies the discussion in this paper is in two parts: (a) What supplies should be carried in a ship, and (b) What provisions should be made for resupply of the ship. The scope of the paper prevents full consideration of the problem, but enough will be covered to prove that there is a difference between the "book" problem and the "real-life" problem.

Professor Huston mentioned "change." It is believed that he had in mind changes which were inevitable because of natural progress. These are bad enough, but when they are combined with those changes which are unnecessary and which seem to stem, at least in part, from personnel turnover in the military administrative hierarchy, a situation is created which is very difficult to describe and to discuss.

The military establishment is a big organization. It is tremendous. There is constant personnel turnover at all levels. The solution for the apparently simple problem which we have undertaken to discuss may be influenced by the following levels in the hierarchy:

1. Congress—goaded at times by the press
2. The President
3. The Secretary of Defense and his civilian assistants
4. The Department of Defense, civilian element
5. The Joint Chiefs of Staff
6. The Secretary of the Navy and his civilian assistants

7. The Navy Department, civilian element
8. The Chief of Naval Operations
9. The Bureaus of the Navy Department
10. Naval Shipyards and other industrial shore activities
11. Supply-Demand Control Points
12. Supply Centers and Supply Depots
13. Fleet commanders
14. Type commanders
15. Squadron commanders
16. Division commanders
17. Ship Captains
18. Heads of departments

No attempt has been made to list these in linear sequence because there is no clean-cut one. Command or control from one level to the other depends very largely on coordination. The evidence in the Navy indicates that good coordination exists only in the fleet; and at the point where it is truly important—within the ship—it is as nearly perfect as necessary.

One simple way of stating the difficulty is by saying that there is too much personal intelligence, too many high-grade people in the vast hierarchy, and too many changes in each position. As each individual accepts or is given a position, he interprets a part of his job to be the correction of things which are wrong. He has a limited time in which to make the corrections. He finds it very easy to find something which needs correcting in the area which he supposes is in his domain, and he starts corrective action at once.

He probably finds something wrong at a low level, and, without taking time to consider what actions are already underway or about to get underway at intermediate levels, he, in effect, starts from scratch. The result is that before any corrective procedure has actually been in effect long enough to be evaluated, another new one is on the way.

Even more common is to note something correctible in the current book procedure and to institute changes in this, so that, at the working level, procedures are never given a chance to stabilize. Therefore, changes upon changes are so frequent that the instructions are never wholly current; and, because of the differences in the time that it takes for an instruction to reach all points at which it is aimed, no two ships are ever acting under the same instructions and doing the same thing.

This is in addition to the effect of self-determination and interpretation. No instructions are so perfect that they will mean exactly the same thing to every receiver. Even if the intent is very clear, the translation of this into mechanics can usually be done in many ways. Perhaps the situation, which is based on whatever past instructions have been interpreted and assimilated, differs from the one which the preparer of the instructions had in mind, so the local interpreter finds it necessary to tailor the new to the old. The old manual-of-arms command, "Fall out and Fall in Again," works fine for a squad or platoon but can hardly apply to administrative procedures, which are in constant operation and cover the entire globe. Since the rapid reorientation is not practicable, it would help the military if every change could be thoroughly thought out before promulgation and afterwards be given time to get effected and settled down before another change is issued.

Since this is asking for more than can be expected from the military, the research empiricist in Naval logistics must take on another task. This is to learn as much as possible

about the real situation at any point from which he collects data and to consider what the real situation will be at any point to which his resultant recommendations will apply. It would require volumes to describe this business in full detail. Even if all of the details which the Logistics Research Project could reveal were reported, there is no assurance that these would cover the desired field. At no time has the Logistics Research Project been in contact with more than ten percent of Naval ships and activities. The one thing that is positively known is that nothing learned from one activity is applicable, with assurance, to any other. The search to find valuable information of broad application from sample data, such as the Logistics Research Project provides, will necessarily be a long and complicated one, but the mass of detail should increase the value of any answers which may be found.

Probably one of the simplest ways to present the scope of the problem is to consider the possible permutations and combinations of only some of the major variables which have to be considered. The requirement is to find a combination of all such variables that would fit each of more than 800 ships. These are:

1. Does the ship carry a supply officer?
2. Are the repair parts binned and drawered? Two possibilities for each of:
   (a) Machinery and electrical
   (b) Electronics
   (c) Ordnance
   (d) Aeronautics
   (e) G-Cog repair
3. Are the bins and drawers under lock and key or on the honor system?
4. Are repair parts centrally controlled? Two possibilities for each of parts (a) through (e) listed under 2 (above).
5. Are General stores (other than G-Cog repair parts) centrally controlled?
6. What allowance list (s) does the ship have?
   (a) COSAL - complete
   (b) SAL or RIAL Machinery and Electrical
   (c) ERPAL - Electronics
   (d) VOCAL - Ordnance
   (e) Aviation
7. For each factor under 6, how current is the list?
8. What is the type commander's policy in regard to in-excess requisitions?
9. What is the ship's policy in regard to these?
10. Where there are centrally controlled storerooms, what is the ship's policy in regard to nonstocked items?
11. To nonstandard items?
12. Has the ship had a supply overhaul and how thorough was it?
   (a) General stores not in Supply Officer's custody
   (b) Machinery and electrical repair parts
   (c) Electronics repair parts
   (d) Ordnance repair parts
   (e) Aviation repair parts
   (f) Equipage
13. What is the ship's custody-receipt practice?

14. What is the ship's practice in regard to allowance-list repairable repair parts?

15. What is the ship's practice in regard to maintaining repair-part inventory records?

16. What is the ship's practice in regard to maintaining machinery histories?

(a) Machinery and electrical
(b) Electronics
(c) Ordnance
(d) Aviation
(e) Equipage

17. What is the type commander's policy in regard to ship's access to scrap and salvage material?

18. What is the ship's practice in regard to this?

19. What is the inventory of repair parts of any category which are in the ship outside of the controlled storerooms?

The list above is rather long, but it does not exhaust the possibilities. Some readers will certainly question the relevancy of certain of these things to a discussion of empirical research. But if each is tested against the following criteria, it is believed that relevancy will be proved:

1. Can the answer affect the nature of data which might be collected to establish empiricism?

2. Can the answer affect the meaning of the data which might be collected?

3. Can the answer possibly color the application of a solution which might be proposed as a result of empirical research?

In summary, if it is assumed that a researcher in Naval logistics is trying to establish an empirical foundation from which he can obtain historical information (data) and that he expects to build models for proposing future improvements; he is cautioned to look to the sources of his data. He should not operate from a gilded tower, assuming that his data are as the "book" would describe them nor that the domain which he hopes to improve is as described in the "book." He must insist on having FACTS.

* * *

PANEL 4
UTILITY AND MILITARY WORTH

MILITARY WORTH AND SYSTEMS DEVELOPMENT

Captain P. L. Folsom, USN

Office of Chief of Naval Operations

This is an unclassified treatment of some problems in military worth which arise in the development of large systems. Recent studies in measures of effectiveness and value theory have been used to define this part of the discussion. Although the discussion will have points in common with many "large" systems, the model used in this discussion is a typical real-time digital data system for executive use. Factors and constraints such as fund limits, pricing, service life, lead time and environment are treated. A typical practical attack on the problem of measuring the military worth of such a system is included. Most systems in this class have not yet accumulated sufficient life in service for measurement and determination of results of plans to "maximize" their military worth. Accordingly, this part of the discussion is restricted to some suggestions concerning defining and measuring military worth of these systems.

MILITARY WORTH

Status of Concept and Theory

Notions of utility and worth, or value, of human actions and material perquisites have attracted the attention of philosophers and economists in increasing measure since the dawn of the industrial era. These notions have arisen, in part, in an attempt to cope with formidable problems, both social and technical, arising in an expanding industrial society. Utility has been defined in a general way as satisfaction of human wants based on a criterion of the greatest good for the greatest number. The ideas of worth, utility, and value have proven difficult to formalize because they are not "absolute," because they have not yielded to a single scale of measure, because of countless possibilities of "exchange" or "trade-off" in measures of value which vary with circumstances, and because it has not been possible in most cases to dissociate moral and social "values" from material values. None the less, scientists are active in studies of worth and value theory but in general their results are still in the initial stages of being transformed into practical tools for executives.

Military Worth

The attention of an important segment of the scientific world has been attracted, more recently, to questions of military worth, which may be conceived as a subdivision of the more general field of worth and value. Interest has been sharpened through the emergence of appropriate analytical tools and the identification of problems which appear to submit, with realism, to formal mathematical treatment. Theory of economic behavior, games, operations analysis, conflict simulation and war-gaming are examples of these new tools. In one way or another these tools promise aid to military or civilian decision-making. The potential assistance from

a set of "pay-offs" or outcomes over a matrix of strategies (or courses of action), as an aid to improving decisions, has become particularly attractive. Here, it is hoped, one may be able to select a strategy or a combination of strategies "best" suited to his ends on the basis of a set of relative "value-graded" predicted outcomes. To date, practical application has proven difficult because military conflict is not deterministic, because military conflict is, more often than not, a "non-zero-sum" contest, because information held by each contestant about his opponent is incomplete and uncertain, and because of the great difficulty in sorting out and handling mixtures of quantitative and qualitative factors. In the war-gaming of certain tactical conflict problems of carefully specified scope, and in the advancement and use of measures of effectiveness, substantial progress has been made. Consideration of problems of broader scope has generated much interest in conflict simulation, with multiple plays and stochastic controls. At present, the tools for dealing with problems of military worth are in the early formative stage and are, at best, of restricted use to executives and decision-makers.

Historical Treatment

The Pyrrhic victory is classic in describing those battles and wars in which immediate military success was attained at too great cost. It is essentially an exercise in "military worth." Through the ages, historians and sociologists have noted the cost and consequences on the fortunes of nations of military establishments and military ventures. These types of analysis enjoyed continuing relevance in earlier times of either relative simplicity and relative stability or of only gradually changing kinds of military capabilities. Times have changed, and the forward surge of military technology has greatly complicated and modified the field for analyses of this kind. Much greater shares of resources are being devoted to providing the "quantum jump" in effectiveness of the individual weapon system. Logistics has burgeoned as the economic brother-in-arms of the combat force. The rapid advances in military technology all too often relegate to early oblivion military systems and capabilities, which only yesterday were dramatic achievements at the very frontiers of contemporary technology. The executive who is supervising system development must, therefore, weigh value and consequences and project the "military worth" of his product, as never before. In this circumstance, the broad historical or social analysis provides him with a framework for adjusting his perspective, but offers little of practical value in aiding his everyday decision making.

Concept for Practical Application

As a consequence of these conditions, the field of research and development, particularly systems development, has become a "battleground" in which a definite atmosphere of battle and conflict, in terms of system potentials, has arrived to further complicate the already difficult technical problems. Expensive undertakings face a higher risk of premature obsolescence. Potential future capabilities of the new system must be considered in value-relationship to current capabilities. Constraints of money, effort, and resources loom ever more important and influential. Current and future requirements are fluid and changeable, and judgements must be made as to what present capabilities to retain and what to discard. Hand-in hand with these problems is the "strategy" aspect of the system development where one desires to obtain a "get there firstest with the mostest system" if he can and, in the inevitable reaction, to outprice his opponent. This paper will not attempt to provide general solutions to these formidable problems. Rather, it will record some attempts to deal with them in a practical way. In this effort recent work of Dr. Douglas Brooks [1] and Dr. Nicholas Smith [2] have been of

much assistance by focusing attention on new ways for appreciating and perhaps dealing with these problems. In order that his paper might remain unclassified, specific references to active development projects and characteristics of particular systems have been omitted.

EXECUTIVE SYSTEMS

Model and Definition

The system selected as the "model" for this discussion is the large-scale digital data-processing system appearing in ever increasing numbers for executive use. In the absence of a generally agreed definition, this class of systems is termed "executive systems." In this broad class of systems the military versions are operational direction systems or functional systems for particular applications, such as logistics. Their counterparts in the business world embrace a wide variety of accounting, inventory, personnel, administrative and computing systems all aimed at saving labor and providing better and quicker information for executive decision-making and management.

System Configuration

These systems are taking the form of central computing elements which are connected to communication circuits of the teletype variety and to displays which show information in symbolic or alphabetic-numeric form or which provide printed data. In these configurations, the computing elements act as calculators and/or as information-sorting, storing, organizing, switching, and routing devices. A variety of card, tape, typewriting, and ancillary storage equipment rounds out the typical system. The characteristics of the particular system and its configuration must conform to the structure, processes, and procedures of the executive organization which it is to serve. Allowing for individual differences between actual executive organizations and the management procedures, there appears to be an underlying logic or ordering of information-processing from raw data to refined aggregates of information which provide the required aid in decision making and management of operations. This ordering of functions appears to be common to all. These systems draw upon known and new techniques, and designing them is a technical art of assembling the proper components to implement ordered executive information-processing procedures into a useable system. As yet, there is no comprehensive and unified body of theory, no general cybernetics as it were, to guide complete system design. Rather, the design is the sum of contributions from many fields of technology. The use of the general-purpose stored-program computer as the main control element of the executive system for operations in real time represents a new application of this device. In this application the program of instructions for the computer establishes the "mode" of operation of the entire system, as well as its own calculating and switching operations. This new application of computers generates a further complication into the technique of computer programming.

Role and Value of Executive Systems

Executive systems, of themselves, do not attain the concrete ends of corporations and enterprises. Their value must be regarded in terms of their contribution—a partial "differential" contribution as it were—to the success of the undertaking and the results obtained. In some cases this contribution is unique and indispensable. That is, it provides capabilities within relevant time frames which could not be duplicated by purely human organization and procedures. In this role, data can be prepared or calculated by executive systems which

would otherwise be out of the question for reasons of excessive cost, time, staff, or demand for minute supervision. In other cases, the value of the system rests on a favorable comparison with other alternatives for executive management in terms of a common measure, such as economy or efficiency. In still other cases, its value may be assessed by its ability to assist the enterprise to attain comparable results, with less implementing force or expenditure of resources than would otherwise be required, or to maximize the output of a stated set of operating resources. These considerations point up the necessity for correct value judgments of enormous consequence at all stages of system development, if the great effort needed to construct them is to avoid producing an inappropriate or irrelevant marvel of automation. There is an urgent need for practical tools which will permit dealing more effectively with those formidable worth and value problems.

DEVELOPMENT PROCESSES AND VALUE PROBLEMS

Sequence

The process of development of a large executive system is a logical sequence of steps starting with the concept of a system (a) to meet a stated requirement or (b) to offer a significant improvement in capability, and ending with a prototype which is suitable for operational use. In this process, the system passes through the main phases of research, development, test, and evaluation. These phases are broken down into more detailed steps, usually starting with long-range objectives. Then, following in order, come research and development directives, a statement of the operational requirement, and a technical development plan drawn up by the developing agency to provide the "how" of the development. Next comes a statement of the detailed characteristics of the developmental system. Construction of the developmental system is followed by technical tests and a technical and operational evaluation. All of this effort is undertaken to provide a system suitable for customer- or service-use. Upon acceptance, the system passes into service procurement and operational stages which lie outside the strict bounds of the research and development domain. Many supporting actions, such as contracting, providing of laboratory and test facilities, providing of simulation and test equipment, funding and accounting, supervision and progress analyzing, and the staffing of the whole undertaking with scientific, technical and military personnel must be undertaken. In an effort to telescope the development and to shorten the initial procurement lead time (currently about seven to ten years) a parallel operation—the training, procurement, installation, and the planning of operational logistic support—is, in some cases, started at an earlier date (during the first or second year of the program). This acceleration requires that development equipment prove out on schedule, that the same equipment be used for both technical and operational development, and that many other actions be interlocked without conflict into an increasingly complex time schedule.

Sources

There are two main sources of system development projects. The first is an attempt to satisfy an operational requirement. Here, present deficiencies or future needs in combat capability are identified as requiring correction through development of a new capability. The second is the development of new techniques which give promise either of meeting known requirements or of providing new capabilities of manifest worth and importance. This latter channel of development may be opened up either as a by-product of general scientific exploration or as a product

of massive exploration of particular fields of technology. Here duplication of effort is often disregarded, and risk is accepted at the outset that much effort may produce no useable results. The production of useable results often is accepted as evidence that the whole venture was worthwhile. Largely negative results may well settle questions of how much further effort is warranted in ploughing any particular field. These methods may be acceptable, in conditions of great tension and urgency, and may also be acceptable where money, availability of technical staff, or availabilities of facilities and equipment are not the controlling limitations. The existence of constraints, particularly limits in funds, is a condition which poses sharp questions of value and military worth in respect to total cost, rate of funding, proportions and shares of resources in relation to other competing demands, the capabilities that will be produced by the effort, and many others. Since development under constraints is becoming more prevalent, it is increasingly clear that reliable estimates of just what the development is producing in improved military capability is needed at the outset of development, and that the best value judgments are required from the very outset of the venture, if the venture is to come up to expectations in terms of useable results for the effort expended.

Requirements

At the outset of this discussion of requirements, note must be taken that the word "requirement"—especially in military usage—connotes a solid, factual, justifiable, indisputable, and overriding need derived by straight-forward reasoning and culminating in precise calculation of kind and number. This conception of military requirements draws a deceptively simple cloak over perhaps the major contemporary problem in executive system development. For it is at this point in the sequence that initial estimates of technical feasibility and outlines of systems concepts will have been completed, and detailed reference to questions of military value and worth and feasibility of the undertaking must be made. In dealing with requirements for systems, there has grown up the practice of stating the requirement, in operational terms, as a desired future combat or operating capability. Here the "measure of effectiveness" provides significant assistance. At this point, the new proposals of Dr. Brooks concerning methods of specifying measure-countermeasure interactions and measures of merit, which are independent of precise details of system performance, seem especially pertinent. In applying these measures, the hunt for measure-countermeasures dead-ends and the determination of the versatility of the system offers immediate aid in grappling with problems of system worth at the "requirements" stage. Given these two factors, measures of effort can be applied to narrow still further the factors which must be considered and to give better assurance that military worth is receiving appropriate consideration early enough in the development sequence. In formulating requirements, the major factors of: "capability in kind" or functional requirements, "capability in numbers" or force requirements, the time frame when these capabilities are needed, and the measures to offset uncertainty of prediction, are used for guidance.

Capability in Kind

The capability in kind is the set of functional characteristics and properties which the system must possess in order to satisfy the expected operational need and conditions, and its minimal manifestation is the existence of a prototype with these characteristics and properties. The prototype is customarily used for experimentation to evaluate its performance. The results provide guidance concerning important judgments and "value choices" which arise

later in the development program. Practical questions which must be dealt with are: the most economical division of maintenance functions between automatic features and technician-tasks; the justification of cut-off point in the face of rising incremental and total-proportion costs in the final approach to complete satisfaction of the requirement; simplification and production engineering substitutes to reduce cost; estimates of suitability to the operating environment, and estimates of long-term logistic supporting costs must be dealt with. In examining the question of "by-products" versatility, and adjustments in design which may be justifiable by providing additional versatility, extensive analysis, through comparisons in terms of measures of effort, is appropriate. Finally, some account must be taken of the correlation between capability in kind and capability in numbers.

Capability in Numbers

With a suitable prototype in hand, the numerical or "force" requirement is an attempt to say how many units of the system are needed. Strategic and tactical considerations, such as the strength, deployment, mobility, and combat characteristics of our own and enemy forces, and the character of possible combat in specified time frames, exert strong influence on the numerical requirement. Practically, working estimates of ranges and strength of force compositions are used for guidance. These estimates lead to sets of finite numbers and set the stage for consideration of the correlation between the capability of the individual system and the numbers required. Future situations are hypothesized to provide a framework for analysis of effectiveness. Mathematical models are constructed in which—among the many parameters—are included numbers, the geometry of employment, and the probability of success of a unit system in a single trial. Using an over-all measure of effectiveness (e.g., the expected number of successful attacking missions for defensive or offensive systems), different mixes of numbers and unit effectiveness may be examined. The alternative may then be priced out and a choice made concerning the effectiveness, performance, and quality of the unit system. In an associated sense, the question of operational reliability as a property of the system, in relation to numerical requirements, opens up value considerations. Indexes for reliability have been derived which permit comparisons of reliability among systems during stated periods of service history. These are statistical and may or may not be reliable in tackling predictions of reliability in new systems, or in trying to determine how much duplication or redundancy is worthwhile in counteracting unreliability. In certain cases of executive system development, 100 percent duplication for the sole purpose of insurance against unreliability has been provided, and, unless the reliability of the single system is quite high, there still remains a significant chance that total outage will occur. Here value-comparison tools are badly needed which will handle unreliability questions, arising from both internal and external sources, in approaches to the goal of one hundred percent reliability.

CONTRIBUTION OF THE EXECUTIVE SYSTEM

At this point it is convenient to demonstrate the possible contribution of an executive system by a military example. Let us suppose that we have six weapons and they are to engage six targets. We have the obvious options of operating the weapons to: (a) engage independently, (b) engage independently within pairs, (c) engage independently in threes, or (d) assign each weapon to engage a separate and distinct target. By using a formula given by Clayton J. Thomas [3], we can compute the probability that at least N targets (N takes values 1 through 6) are engaged

for each of the above options. It is assumed that each weapon will engage and that, in independent assignment, any weapon is equally likely to engage any target. These values are given in the following table:

TABLE 1

Approximate Probability That At Least N Targets Are Engaged

N=Method of Operation	1	2	3	4	5	6
Independent Pairs	1.00	.99	.98	.76	.24	.02
	1.00	1.00	.99	.84	.35	.05
Threes	1.00	1.00	.99	.88	.50	.13
Explicit Assignment	1.00	1.00	1.00	1.00	1.00	1.00
9 Weapons Independent	1.00	1.00	.99	.97	.67	.20

Note that for comparison we have added a line showing the engagement probability if we were to employ nine weapons independently against the same six targets. From the data we may judge that increasingly explicit coordination raises the engagement probability only relatively slowly and that we make dramatic gains only if complete coordination is provided. Here we would probably be justified in providing executive system facilities only if they were capable of effecting complete coordination of engagement. The value judgement comes next: how much should we pay for this capability? If the cost of weapons is very small in comparison to the executive system, there is doubt that the executive system should be provided. If, on the other hand, a weapon is several times as expensive as an executive system and if, further, we are limited by cost, transportability, or some other constraint in the number of weapons available, a proficient executive system would seem to be very strongly justified. Conversely, the common temptation to buy "just one more weapon," at the expense of the ability to coordinate them all, is a false economy.

Time Frame

The time frame in the future, during which the capability of the new system is required, is exerting an ever-increasing influence on the worth of the system. A well-known, alarming effect of rapid progress in military technology is the appearance of dramatic new capabilities at intervals short in comparison to system development lead times, of the order of seven to ten years, which we have experienced since World War II. More often then not, these new capabilities render other recent and costly system obsolete. It can be argued that some processes are more susceptible of speed-up than others or that there is really no speed up as such, but rather a better selection of developments within resources, or a quicker response to break-throughs. None the less, there is a major requirement for reliable tools to cope with value analysis incident to ensuring continued viability and effectiveness of systems over an extended period of service life. Inasmuch as the service epoch of the system is preceded by a development period, in which time for development and development resources normally can be used once and only once, this factor exerts great

influence in adjusting the trade-off between capabilities of systems in terms of sophistication (however long it may take) and time available to attain a worthwhile order of improvement by the operational due-date of the system. Here a concept of a "system family" may prove useful, in which a larger concept is implemented in part in an early and realizable schedule, with the arrangements made for follow-on additions during service life. In executive systems—in particular where new applications of the system can be attained through new sequences of instructions, rather than through substantial engineering alteration or even replacement of equipment—a better value in keeping the system "up with the times" as well as more favorable cost comparison may result. A particularly sticky point in the "time frame" problem is the estimate of whether or not the system is likely to be used for combat purposes in the time frame. The difficulty appears to arise from want of measures of value of "the system in being" during the time frame, even if it is never in combat. Time-frame considerations, of which there are more than have been indicated above, in a climate of competitive system development, seem to impart to systems developments the quality of "strategies" over and above their character as engineering enterprises. Here a sort of indirect conflict obtains in which potential capabilities are pitted against one another, often in unfinished form, with a simulacrum of maneuver in an attempt to outwit or out-cost the opponent. Tools for value measurement and aid in decision in this development climate are imperative needs.

Cost and Funding

Threaded throughout the development process are innumerable problems in the area of costs, prices, and funding. Inasmuch as money is a principal constraint on system development, as well as being a ready-made measure for estimating and comparing systems and other competing developments, total costing out of entire programs is a necessity. The cost analysis can be divided into two main parts: the cost of research, development, test, evaluation, procurement, and installation, and the cost of sustained service operations thereafter. The itemized cost of all contributing tasks can be aggregated into major titles, such as:

1. Research, Development, Test, and Evaluation
2. Procurement
3. Operations and Maintenance
4. Manpower and Personnel
5. Facilities and Construction

Funding-time increments, such as fiscal years, are usually imposed for budgetary control and accounting. Throughout the development, it is useful to provide the capability to establish new planning datums and these can be conveniently established at the beginning of funding-time increments. The reestablishment of new datums is made easier if the aggregated accounting of money spent is performed in the same terms as future cost analysis, although in some instances changes to major titles are only approximate and, in some cases, arbitrary. The aggregate costs under a particular title then, are:

$$C\ (\text{title}) = \sum^{\text{year}} C(\text{title, year})$$

and the costs in any year are:

$$C\ (\text{year}) = \sum^{\text{title}} C(\text{title, year})$$

making possible the check:

$$\sum_{\text{year}} C(\text{title}) = \sum_{\text{title}} C(\text{year})$$

which is the total cost to date. The future cost analysis is subject to all of the well-known uncertainties that attend price estimating but, with the added safeguards of applying dollar devaluation and the use of running judgment factors and frequent price-checking and updating, a measure of control can be attained. In order to provide a basis for future costing, projected installation and manning schedules are required. Through the use of planning factors, all costs related to service operation and maintenance can be estimated. With the installation schedule in hand, the total of unit-system years of capability can be projected. If, now, the total cost to date is related to the aggregate of unit-system years of capability, a measure of the running cost situation can be derived. It is:

$$\text{Cost per unit per year of capability} = \frac{\text{Total Cost to Date}}{\text{Aggregate System-Years of Capability}}$$

A trend for any number of years in the future, consisting of the cost per unit per year of capability for each year in the future, can be computed. Variations in the trend are then available at each updating of the cost analysis. In this measure, each system is amortized fully to the year of computation thus avoiding the need for an arbitrary amorization period. With these data in hand, it is possible to make some estimates of annual and long-term costs of executive systems in relation to the budget of the entire enterprise or corporation which they support. Here, two points are of interest: The first is to detect at an early date any tendency of supporting costs during the operational era to rise excessively, and to take remedial action during development. The second is to compare the cost of the direction system in relation to the total cost.

SOME VALUE CRITERIA FOR EXECUTIVE SYSTEMS

Practical Criteria. It is now appropriate to indicate some of the practical criteria which are currently in use to assess the worth or value of executive systems. These are no more than partial answers to complex problems in assessing military worth and will give way to better methods when these become available.

Requirement. This criterion attempts to say whether or not proficient direction and coordination is vital to the enterprise.

Uniqueness. This criterion attempts to assess the worth of executive systems on the ground that information requirements for decision and direction of operations cannot be met by the alternate techniques of human organization aided by special data and communications devices. The allied question of whether or not the system is excessively expensive in relation to the "value" of the operations it directs is not yet satisfactorily answered, even though this criterion may be met.

Economy. In cases where the function of decision and direction may be performed by either a direction system or a human organization with aids, the question arises as to which is not economical in money. Assuming that either can be afforded, the executive system qualifies if it is more economical.

Contribution to Payoff. In this criterion our purpose is to provide a measure of comparison between payoffs for the conditions of poorly-coordinated vs well-coordinated direction.

The foregoing weapon-assignment example is a case in point. It leads to a consideration of shares and proportions of cost of executive systems and entire operations. Measures of reasonable ratios of sustaining cost to cost of development and procurement are not available, and logistics-research assistance is needed to provide them. Furthermore, no reliable figures of reasonable ratios of cost of executive systems to total costs are available, but there are indications that executive-system costs should be small in relation to total cost and that they may be nearing the feasibility limit at a 10 percent to 15 percent level. The provision of reliable guides in this matter is a worthwhile goal for logistics research.

Feasibility. In addition to repeated checks for financial feasibility during the development (and later operation) of the system, feasibility in other features of the system must be repeatedly estimated and evaluated, since failure to prove feasibility in any one of them may render the system worthless for the intended purpose of requiring excessive reconversion of resources at the expense of other programs. Here, number, qualifications, training requirements, and adaptability of personnel must be checked out and proved out in tests. The ability of industry to manufacture and deliver on schedule, and the ability of the engineering establishment to install and check out equipments on prescribed schedules must be reaffirmed frequently.

Versatility. There is a set of cases where the future requirement is uncertain but is expected to become more rigorous. Here it is desirable that the executive system be able to adjust rapidly and economically to new conditions. In a number of these cases the change may be so drastic that the continuance of the enterprise depends largely upon the versatility of the system. If margins for versatility can be incorporated at the outset without exorbitant extra cost, it becomes a candidate under this criterion. In executive systems, such properties as add-on capability; multiple sequence control for ease in adaptation of modes of operation; common arithmetic and logical elements; a common binary language for both calculation and information distribution, and a capability for trade-offs among the controlling information parameters give much promise of providing this versatility. If verified by comparative analysis, executive systems may qualify on the grounds of versatility.

Reliability. In cases where the reliability of executive systems show markedly superior technical reliability and the cost comparisons with other techniques are not unfavorable, the executive system may qualify on the grounds of reliability.

Assessment of the value of executive systems on these, and perhaps other, criteria is not a question of simple suboptimization, but rather a practical attempt to optimize over-all criteria. Military worth might be handled by a typical project manager somewhat along the lines of the System Development Flow Diagram (Figure 1). There has not been sufficient experience with executive systems to date to evaluate the validity of these criteria. They are guidelines at best and will give way to better methods when these become available. In any attempt to provide supervision of systems-development using more reliable, but undoubtedly more sophisticated value-assessment tools, it is all but certain that substantial data automation will be indispensable.

CONCLUSIONS

With the recent development of executive systems as a background, an attempt has been made to show some examples of how value-assessment problems associated with these developments are handled. The outstanding conclusion is that supervisors of these developments are badly in need of improved value-assessment tools. The second conclusion is that, in common

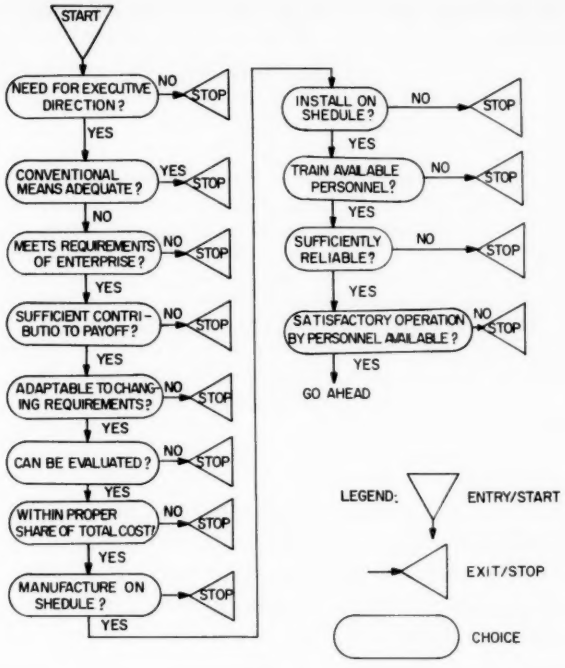


Figure 1 - System development flow diagram

with other developments, the development of executive systems has become not only a complex engineering enterprise, but a battleground of strategies as well. The third conclusion is that, although studies in theory of value and worth are proceeding apace, some early concession to improved empirical tools is warranted. The fourth conclusion is that many of the problems of executive-system development are logistical in nature, and logistic considerations must be fully reflected in new methods for assessing the military worth of these systems. The fifth conclusion is that, whatever tools for assessment of worth are devised, ceaseless reaffirmation during development of the total feasibility of these systems will continue to be a fundamental requirement. The final conclusion is that if improved value-assessment tools are provided by the scientific world, data-automation aids in their application is an almost certain requirement.

REFERENCES

[1] Douglas L. Brooks, "Choice of Pay-offs for Military Operations of the Future," presented at 16th National Meeting of Operations Research Society of America, Pasadena, California, November 11-13, 1959 (Unclassified).

[2] Nicholas M. Smith, Jr., "A Calculus for Ethics: A Theory of the Structure of Value," Parts I and II - Operations Research Office, The Johns Hopkins University (Unclassified).

[3] Clayton J. Thomas, "The Estimation of Bombers Surviving an Air Battle," Operations Analysis Technical Memorandum 49, August 1956 (Unclassified).

* * *

PRIORITY RATINGS IN MORE THAN ONE DIMENSION*

W. G. Mellon

Princeton University

This paper presents certain theoretical and empirical results for problems which involve the assignment of priority ratings or indicators of military worth, where either several of these ratings must be aggregated or the outcome is to be evaluated by a multivalued criterion. The first portion of the paper contains a discussion of the logic of the problem and the axiomatic structure of a desirable solution. The second section includes the results of an experiment in which random demands were processed by running them through a model supply system using several alternative systems of priorities.

INTRODUCTION

These remarks are a by-product of research being carried out by the author at the Econometric Research Program of Princeton University, under the sponsorship of the Office of Naval Research.[1] Over the past two years we have been attempting to lay the theoretical foundations of a theory of priorities. A word of explanation with respect to terminology is in order. By the word "priority" we mean all indicators of relative desirability, no matter what degree of measurement they contain – that is, whether they represent a simple ordering of alternatives, an ordering on the simple or compound distances between alternatives, or a numerical scale which is applied to the alternatives.[2] This definition represents merely a terminological whim – a desire to divorce our discussions from the particular connotations which have sprung up about the word "utility" in economic discussions over the years. "Priorities," while they have, in general, meant only ordinal indicators of desirability, begin their theoretical life here with a clean slate in the way of connotation.

As a small representative sample of our investigation, a particular subsection of the problem involved in the development of priority theory has been chosen. In the remainder of the paper, we shall do the following:

a. Explain the nature of this particular aspect of the general priority problem.

b. Give some of the theoretical considerations involved.

c. Outline some suggestions for the sort of research which seems called for, if such problems are to be correctly handled.

The last item occupies last place as a representation of the author's convictions on the logical relationship between theory and empiricism in such problem-oriented studies as those involved with military supply. This is that the line of development should be from

*The author wishes to thank Mr. Richard Davis for some interesting discussions on this general topic.

[1]Under contract No. Nonr 1858 (16). For further details see [1].

[2]By "numerical" scale we mean uniqueness up to a linear transformation. In questions involving "military worth," it is unlikely that we shall ever deal with the measurement of anything which involves a ratio or absolute zero scale: See [1], Note 1, Appendix 6.

theory to empiricism to reformulation of theory. It is true that, in aspects of decision theory relating to military supply, a great deal of attention has been placed on the development of theory which is perhaps overly elaborate, especially when we remember that in most military supply problems we are seeking improvement rather than optimization — the concept which has come to be known in military logistics as "Acceptability." But, on the other hand, much of the empirical research done in this field seems to have either suffered from certain logical errors or simply to have been done in ignorance of the fact that certain logical problems were involved. It is usually best to accept the guides which theory can provide in defining optimal solutions. At least, this is the basis of our approach.

DISCUSSION

The title of this discussion is given as "Priority Indicators in More Than One Dimension." This is, in fact, rather elaborate nomenclature for a very common, if not easily resolvable, problem. In a number of situations in economic or military-decision theory we find that the "payoffs" from various alternatives are not single values, but a set of vectors; on the other hand, it is necessary to rank these alternatives, the degree of measurement being left unspecified for the moment. Some examples will make the general problem clearer. In the case of assigning naval electronic equipment, which has been discussed in a series of papers in the Naval Research Logistics Quarterly, the authors determined separately the "worth" of a piece of electronic equipment and the importance of ships to which such equipment might be assigned. Then these two measures of worth were combined to give a numerical measure of the "worth" of assigning each particular piece of equipment to each ship. Afterwards, this matrix of coefficients was used in the computation of an optimal assignment problem of the usual sort, where constraints are imposed by the number of each type of equipment which is on hand and/or may be purchased out of a limited budgetary appropriation [2, 3, 4, 5]. The work of Dr. Henry Solomon and his associates represents another variant of the same problem, in that he treats the military worth of a spare part as a combination of the effect of the failure of a part on the component in which it is installed and the effect of a failure of this component on the ability of the ship to perform its assigned mission [6, 7]. To this consideration, we could add the value of the ship's mission to the general strategic plan for its area. Finally, Captain P. L. Folsom has pointed out in his discussion the difficulties in ranking various military control systems, when these systems differ from one another with respect to various criteria of desirability, such as cost, dependability, maintenance, and so on [8].

The general problem in simplest terms is this: given a number of points representing alternatives which exist in a space whose dimensions are the evaluation or appraisal of the alternatives in terms of particular attributes, how are these alternatives to be ranked? Although many of the more interesting problems are those which deal with several dimensions — that is, where the space involved is the positive hyperoctant — for illustrative purposes we shall use a two-dimensional example. Suppose that we have a central supply depot which supports a number of outlying bases. Each of these bases submits, from time to time, requests for material to which they assign a priority classification, which represents the relative value of that material to the particular base. Presumably, the central depot, in turn, is able to appraise the relative value of the activities of each base, again according to a set of priority classifications. On the basis of these two sets of priorities, the central depot must rank the requests in order of importance. The problem is one noted by

Oscar Morgenstern [9]; see also [10]. We omit the discussion of the time dimension, though it is perhaps the most important, since our purpose is purely illustrative.

We may first approach the problem in terms of the degree of measurement which we happen to possess in each of the dimensions. The possible combinations are shown below:

Possible Degrees of Measurement

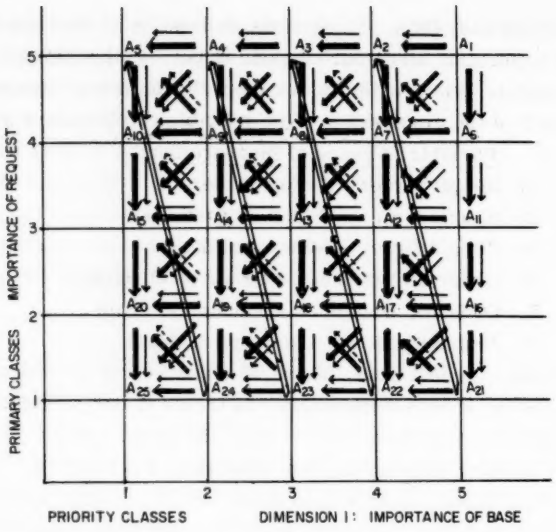
1. Simple Order - Simple Order
2. Simple Order - Order on Distances
3. Simple Order - Numerical Scale
4. Order on Distances - Order on Distances
5. Order on Distances - Numerical Scale
6. Numerical Scale - Numerical Scale

We begin with the first and simplest case, where we possess only an ordering of alternatives according to their properties in both dimensions. In terms of our example, let us suppose that there are five priority classes into which each of the bases can put their requests, and that, in addition, there are five priority classes describing the relative importance of bases as appraised by the central depot. In this case there are 25 possible priority combinations (25 positions in our two-dimensional space) to which any particular request may be assigned, depending on the two priorities attached to it. The problem is to rank these 25 points in order of importance, thus giving us a partial order of importance for the requests for materials which are submitted.[3] The total number of possible orderings is 25! We wish to find criteria which will eliminate at least some of these. Such criteria do exist, and two are discussed below.

A. The first criterion we consider has a desirable feature in that it completely orders all our positions; but it is, in fact, rather too strong to be generally accepted. This is to assume that the ordering of the alternatives is lexicographic. In such orderings, one vector is considered to be greater than another if one particular component of one is greater than the equivalent component of another; the relations between all other components being irrelevant. Only if the two vectors have equal values of the first component, will the determination of which vector is greater rest on the comparison of other components. In terms of our supply example, this would mean that we would decide that one of the dimensions was the really important one — say the relative importance (as given by the priority ranking) of the base to which the material is sent. Therefore, all requests from bases of priority 5 would be processed ahead of all requests from bases of priority 4 — no matter what priorities the bases themselves attached to their requests; and that only if the requests were from bases of equal importance, would the ranking of the alternatives be decided by utilizing the priorities attached to the requests by the bases themselves.

The reader is referred to Figure 1 which indicates how each of the criteria we discuss generates a set of relations between the 25 alternatives shown. It is easily seen that the assumption that the order is lexicographic and that 1 is the dominant vector generates a complete order of the form $a_1 > a_6 > a_{11} > a_{16} > a_{21} > a_2 > a_7 \cdots\cdots$.

[3] We may assume that when a number of requests have identical ratings in both priority dimensions that they are processed on a "first come, first served" basis.



Legend

1. Arrows ➡ and ⇒ indicate the preference relations established by the assumption that the ordering is lexicographic; relations between alternatives which are not shown are ensured through the property of transitivity.

2. Arrows ⇢ indicate the preference relations established by the assumption of Pareto Optimality.

3. Arrows --> indicate the preference relationships established by the simple ordering on the dimensions.

Figure 1 - Orders under alternative criteria

The assumption that we are dealing with a lexicographic ordering is not an acceptable criterion in most cases of military supply. We would not ordinarily say, for example, that a set of ping-pong paddles for an important combatant ship is more important than the engine for a supply ship. In most cases, an infinitesimal movement along one of our dimensional axes, does not outweigh very large movement(s) along the other(s), which the assumption of lexicographicity indicates. This means that we must look for a weaker, but more acceptable criterion of ordering.

B. One criterion exists which is intuitively perfectly acceptable: this is the rather obvious concept known to economists as the Principle of Pareto Optimality.[4] Under this criterion one alternative is considered superior to another if it exceeds the latter in at least one dimension, and is no worse than the second in any other dimension. In terms of our

[4]As a parenthetical comment, we may observe that much of the ground which is being covered by military worth studies is already familiar in other contexts: to economists as the theory of welfare economics, and to psychologists as the theory of psychometrics. In particular, with reference to the former, while the achievements of the "New" Welfare Economics are mostly negative, in that they point out what cannot be said concerning the solution of certain problems, a brief study of these results may at least keep the military-worth analyst from wandering down some well-worn blind alleys.

illustration, more urgent requests from more important bases will take precedence over less urgent requests from less important bases, more important requests from equally important bases take precedence over less urgent requests, and requests of equal urgency take precedence according to the importance of the base from which they originate.

We may refer to Figure 1 to determine the effect of this criterion in establishing an order among our 25 alternatives. From the observation of the light arrows we see that the assumption of Pareto Optimality clearly establishes alternative a_1 as the most highly preferred alternative and a_{25} as the least preferred, and also establishes certain relationships between all the other alternatives. The criterion does not determine a single unique order, as did (A), but does reduce the number of possible orders which meet the criterion to a tiny fraction of the total possible orders. If we use the property of transitivity ($a_1 > a_2 > a_3 = > a_1 > a_3$), then a series of 23 comparisons will be sufficient to always establish a complete order on the 25 positions. The logic of those questions is as follows: we shift to a 3 x 3 example here so as to narrow the space required for the illustration. Let the points to be ordered be as in Figure 2. In the final order which we are seeking we know the final term, i.e., a_1: the second term may be either a_2 or a_4. This is determined either by posing a direct question or imposing an additional decision criterion. Suppose that a_2 is chosen as the second term; then the choice for the third term is either a_3 or a_7. This process generates a tree diagram of the form shown in Figure 3. By following down the diagram and obtaining the relevant information at each step, the complete order on the alternatives is obtained by 7 questions, where we are dealing with 9 alternatives, which enables us to choose 1 of the 42 possible complete orders which obey Pareto Optimality in the 3 x 3 case, and by 23 questions in our 5 x 5 example.

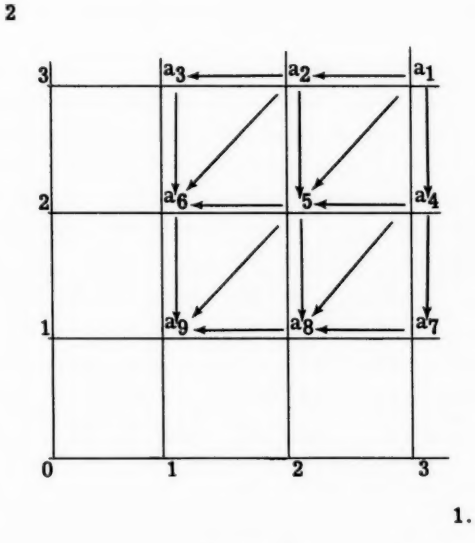


Figure 2

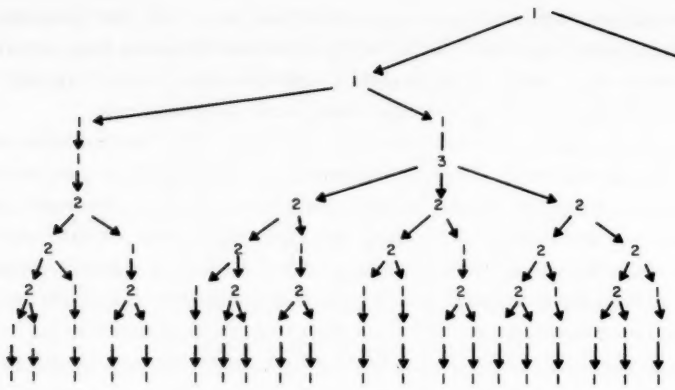


Figure 3 - Ordering tree for 3 x 3 example

As we have noted, there are two possible methods for obtaining the necessary information:

1. By asking qualified personnel to make the comparisons for you.[5] Thus, we might ask a series of questions of the form: "Do you regard a request of priority 4 urgency from a base of priority 3 importance as taking precedence over a request of priority 3 urgency from a base of priority 4 importance?" By setting up the relationship in the form shown in Figure 2, we will be able to ask the relevant, and only the relevant, comparisons and keep the questioning to a minimum.

2. By imposing additional criteria over and above the Pareto Optimality. Suppose, for instance, that we merely order the dimensions themselves – that is that we get an opinion from qualified personnel that, other things being equal, we should consider the importance of the base from which a request for material originates as a somewhat more important consideration than the urgency of the requests themselves. This will provide us with 16 additional pieces of information (shown as - - - - >in Figure 1) plus other information that can be obtained through the application of the principle of transitivity. The maximum number of additional bits of information needed to establish a complete order is then once again reduced.

[5] It might be (undestandably) difficult for the relevant authority to make this sort of decision in vacuo. We have, therefore, designed the following method: A model of the supply system is constructed in the same manner as the one given in the example above. Into this supply system are fed a number of randomly-generated requests for information: certain fixed times for processing a request are built in, as well as the delays which are encountered by requests which are given a lower priority rating. Thus, when a particular method of determining the relative priority of differing requests from the bases has been determined, the model will generate a particular pattern of delivery on these requests. By observing the comparative patterns when different methods of ranking the requests are used, the authority may be in a better position to approve which method he prefers. The method can also be used to appraise whether or not a general change in the priority system—such as altering the number of priority classification—is desirable.

For our 3 x 3 example, there are only 4 complete orders which obey this criterion plus Pareto Optimality: we therefore need to ask only 2 direct comparisons to establish the final complete order. See diagram in Figure 4.

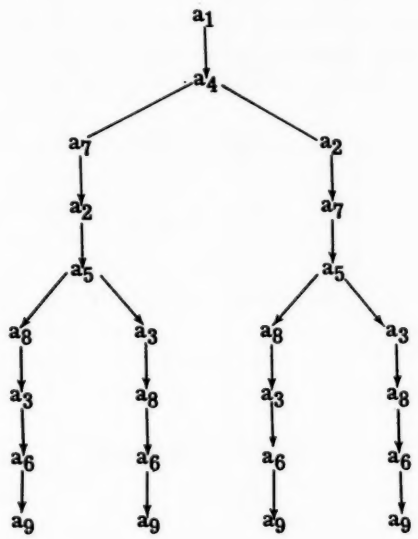
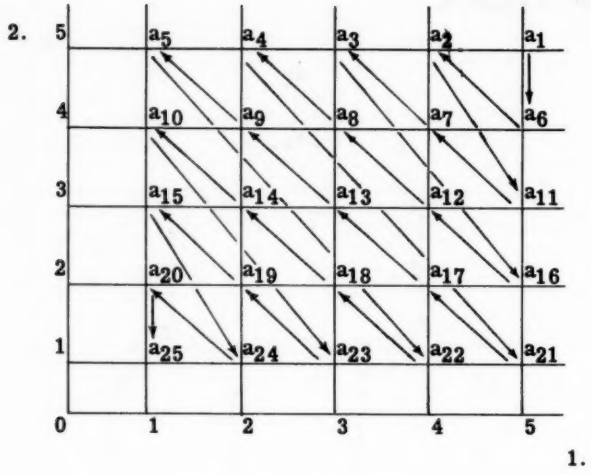


Figure 4

Again, we may either stop at this point and ask for the necessary comparisons directly, or we may impose still further additional criteria.[6] The next logical criterion would be as follows: an alternative is said to be better than another if it exceeds the latter in one dimension by at least one priority grouping more than it is exceeded by the latter in the other dimension. In terms of our example, this means that an alternative consisting of a request of urgency 4 from a base of importance 3 would be preferred to a request of urgency 2 from a base of importance 4. When this criterion is used together with ordering the dimension, the transitivity, and the Pareto Optimality, we obtain a complete order represented by Figure 5 (assuming dimension 1 > 2).[7] It should be emphasized that criteria like the last one introduced may sound reasonable, but there is no reason why they should hold in particular

[6] It should be realized that just as the number of pieces of information necessary to yield a complete order increases rapidly as the number of dimensions and the degrees of measurement in each dimension increases, so does the strength of the criteria needed to obtain a complete order on the alternatives.

[7] This is equivalent to the following criterion: let the alternatives be ordered by their Cartesian distance from the origin. Where ties occur, these are broken by the ordering on the dimensions.



(only final order shown)

Figure 5

situations, for they always imply an assumption of constancy of relation which certainly need not of necessity hold throughout the whole system of alternatives.[8]

For those who prefer numbers to order relations and for simplification of the computations where we have a number of dimensions, we may employ the following method. To

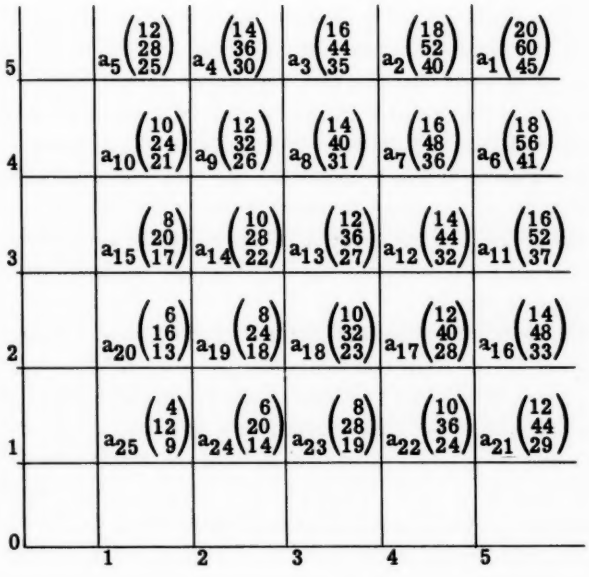
[8] The main alternative method of approach to the problem would be to apply the reasoning of what has come to be called "decision theory." Thus, if we have a number of alternatives and a number of dimensions in which they are to be ranked, with five possible rankings for each alternative in each dimension ranging from 1 (low) to 5 (high), we have a tableau of the form given below. Again, we wish to rank the alternatives themselves. A number of criteria have been suggested for this purpose, including those of Laplace, Wald, Savage, and Hurwicz. As these have all been given in the context of uncertainty, a certain amount of adaptation is necessary to fit this reasoning to the problem of how to rank the rows, when they are treated as alternative certainties in several dimensions. This question is discussed at length in [1].

	Dimensions				
Alternatives	1	2	3	4	5
1	1	2	3	4	5
2	4	5	3	2	1
3	5	4	1	1	4
4	3	2	5	1	2
5	2	2	2	3	3
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.

each of our five priority categories in each of the dimensions, we assign the following numbers arbitrarily.[9] Suppose that we are dealing with a problem which involves a number of dimensions, and that we are able to rank these dimensions in order of importance (i.e., low cost means less than easy maintenance, which means less than dependability). To each of the dimensions we might assign the following numbers, arbitrarily, but in such a way that these weights, when multiplied times the priority weights, generate a series of numbers whose order corresponds to the criterion of ordering which we have adopted.

Priority	Score
1	2
2	4
3	6
4	8
5	10

2



1.

Figure 6

TABLE 1

Dimensions	Scores		
1	2	3	4
1	2	4	2.5
2	2	2	2.0
3	2	.	.
4	2	.	.
.	.	.	.
.	.	.	.
.	.		

If we use the weights in column 2 of Table 1 (which are in reality no weights at all), the numbers generated (top figure in each bracket in Figure 6) order the points according to Pareto Optimality. By using the weights in column 3 of Table 1, we generate a series of numbers

[9] Any set of numbers which represent a monotone transformation would do as well, unless one wished to suggest that the priority classes are in some sense equally spaced.

(middle figure in each bracket in Figure 6) which obey Pareto Optimality + ordering on the dimensions (1 > 2) + transitivity. By using the weights in column 4 of Table 1, we can generate a series (bottom figure in each bracket in Figure 6) which obeys these criteria plus the additional criterion suggested in Figure 3. These "scales" obey the criteria, but only in the sense that they are unique up to a monotone transformation: this property must always be kept in mind, especially if one uses the scales as if they had "true" numerical properties. To do so is not "wrong," but the appearance of accuracy in the result is, of course, largely fictitious. If only ordering data is available, but numerical computations must be performed, scales of this sort may be used with due caution.

We now move to a consideration of the next degree of measurement which we may happen to possess in each of the dimensions: this we have called "order-on-distance."[10] It may be that, through the questioning of competent personnel, we are able to obtain a measure of the relative distances between the attributes of alternatives in particular dimensions: for example, that the durability of equipment A is greater than the durability of equipment B, which is greater than the durability of equipment C; but also, for example, that the difference in durability between A and B is greater than that between B and C. Suppose that we consider the ranking of alternatives A, B, C, and D in a particular dimension. If we fix the interval and unit of measure – that is, assign the lowest alternative a value of, say, zero and the highest a value of one, then given the order on the distances between the alternatives, the numerical value which may be assigned to any alternative is confined within a certain range. The greater number of alternatives, the smaller the possible range. At the limit where we have an infinite number of alternatives, this order on the distances will generate a numerical scale unique to a linear transformation. For a proof of this proposition, see [12]. In order to conserve space, we may now skip a consideration of many intermediate problems, and, after a slight digression, go right on to a consideration of the question of how to obtain a complete order on alternatives when the scales in all the dimensions have a numerical property (No. 6, under Possible Degrees of Measurement on page 3).

In the context of military worth, such information is almost certainly never available to the analyst – that is, we cannot expect anyone to tell us that a Mark IV electronic equipment is 7.234 times as valuable as a Mark III model. Therefore, any numerical properties which we attribute to our scales is an addition to the information really on hand, and is, in one sense, fictitious. We have previously seen how a numerical system consistent with certain criteria on simple orders can be developed. If the information is in the form of orders-on-distance, then we may take the midpoint of the numerical interval determined by such information when the total interval is bounded, as the scale which is to be attached to each alternative. This is the most obvious choice, and, if we wish to conceive of a "true value" of the alternative (which, due to our ignorance, we can conceive as lying anywhere within the numerical range determined by the order on the distance), this choice minimizes the total possible deviation of our scale from the "true" scale.

[10] Such scales are also called "metric." For a discussion of their properties see Siegel [11]. Professor Siegel uses a simplified version of the von Neumann-Morgenstern method for obtaining a numerical utility to obtain his information on the distance between alternatives, rather than a direct question method.

Let us then suppose that by this method – or by some other, i.e., the von Neumann-Morgenstern utility method[11] – we possess a scale that is unique to a linear transformation which ranks the various components of a set of multidimensional alternatives. Once again we face the general problem of priority indicators in more than one dimension: how are these alternatives to be ranked by a single-valued numerical scale? We consider here three possible types of answers to this question:

The first of these is to fall back on our assumption of lexicographic ordering. Suppose we should have several alternatives, as represented in Figure 7, with their numerical rankings in both dimensions. Since the scales in both axes are continuous, there are an infinite number of possible alternative parts; but the 25 chosen will illustrate any point, and will also put the discussion on a comparable basis with those remarks which preceded it. Assuming 1 as the dominant dimension, the a- points will be ranked as before: for every one of the a points which has the same second component, we will know precisely by how much one differs from the other; and for every one of the a points which has the same first component, we will know precisely by how much it differs from the others. The order for these is given in Figure 1. Even if we were to accept the absolute dominating importance which the ordering implies, it can be readily seen that the ordering does not, and cannot, establish a continuous meaningful numerical scale for the alternatives – meaningful in the sense that we can say that a_{13} is .45 better than a_{20}, see [14]. The only situation in which such an ordering would be very helpful would be one in which we were trying to select a single best alternative from a group and to use the value attached to it for computational purposes. Thus, if our alternatives were, say, a_{15}, a_{14}, a_{13}, a_8, and a_3; we could immediately discard a_{15}, a_{14}, and a_{13} as dominated by the others in the first component; and then pick a_3 as superior since it exceeds the others in the less important component 2. Such multidimensional utilities could thus find obvious uses in game theory, where we are seeking a best strategy, or in linear programming, where we are seeking a best point [15, 16], but they would be of little use in such problems as selecting an optimal set of alternatives to be acquired with limited resources.

The second method for resolving the multidimensional priority problem is the one which we may call "conversion." Here a factor is established by which measurement along one axis is converted into measurement along another. A simple example will make this clearer. Let us suppose that Congress wishes to establish a set of rules which will give a numerical ranking according to which applicants for government jobs will be rated. In establishing this rating, they decide to take two factors into account: (1) the suitability of applicants for the position for which they are applying, as measured by their numerical scores (from 0 to 100) on a standard Civil Service examination, and (2) the applicants' length of military service, measured in months. Thus, each applicant can be represented as a point in a two-dimensional space where both axes have a numerical scale, but where the two scales are not in comparable terms.

Suppose, then, that we adopt the following rule: add 101 points to the score of every veteran who has served at least one month and an additional point for every additional month of service. Thus, we can easily see that this rule will give us a continuous scale in terms of test scores in which all veterans rank above all nonveterans, the important feature here being

[11] The difficulty with this method is, of course, the very large number of questions which must be asked before we can establish the numerical scale by which a person values the set of alternatives. To make the N-M method operational, we might search for additional plausible conditions which would enable us to deduce the entire scale from a limited number of questions. For an interesting attempt in this direction, see [13].

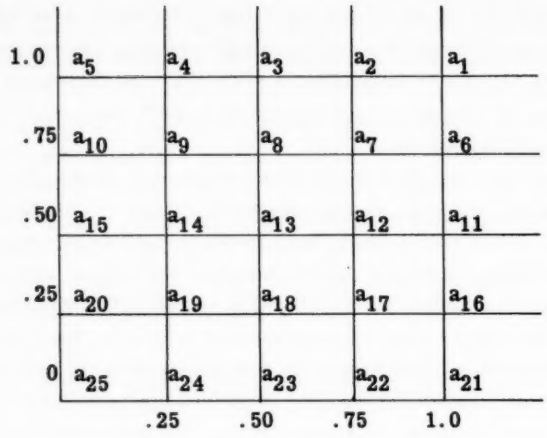


Figure 7

the degree of military service (for equal periods of service the deciding factor is to be the test score). The ordering is then also lexicographic. This extreme preference may be weakened by assigning a smaller weight to military service. Thus, we might award a flat 10 points to all veterans, one point for each month's additional service and so on.

A method of this sort may be used to resolve a number of multidimensional priority problems, especially when we are seeking a convenient rule-of-thumb. On the other hand, it does involve a very strong assumption which is probably inapplicable to most situations: that the contribution of the vector, which is converted into the other to the total numerical score of each position, is constant – in our example, that one month's service is always equal to 1 point of test score, no matter what the score may be.

The third method which may be used to resolve this problem is the method of "combination," in which the scores of an alternative along all axes are "combined" – that is, added and/or multiplied together to give a total score for each alternative. This is, of course, adding apples and peaches since the measurements are not in comparable terms; but if we are careful to always remember the dubious numerical nature of the scales with which we are dealing, we will find the method as satisfactory as any of the others we have mentioned. The general choice in this instance is between addition and multiplication. If properly used, both of these methods will usually preserve the property of Pareto Optimality, and will indicate whether each alternative is $>$, is equal to, or $<$ than any other. In general, the choice of whether one uses addition or multiplication would depend on the particular problem at hand.[12] Thus, if your lowest ranking alternative was, in fact, completely valueless in one dimension (as in the first example in this discussion: if one of the outlying bases were of no positive value to the military effort), then it would make sense to assign this alternative the

[12] If we possess the final figures themselves (in the assignment problem, for example, the value of item X in position Z), then we can show that this figure can be broken up into the sum of two values which represent the contribution of each dimension to the total: See [17].

value of zero and to use the rule of multiplication. In our example this would be equivalent to stating that a request from an activity is worthless, no matter what is degree of urgency. On the other hand, it would be obviously absurd to give the lowest degree of urgency a value of zero, since this would assert that the lowest ranking request from the highest valued base (in terms of importance) should be treated exactly the same as the lowest ranked request from the lowest ranked base. This difficulty can be easily avoided with the following procedure: let the numerical scale, which is assigned by the outlying bases to the relative urgency of a request, run from 1 to 2 and the relative importance of the base from which the request originates run from 0 to 1. All requests from the worthless base will be treated as worthless, all lowest value requests will be ordered according to the importance of the base from which they originate, and the principle of Pareto Optimality will be observed. If the lowest ranked base is, on the other hand, not worthless, but simply least important, we may have both scales run from 1 to 2.[13] To re-emphasize what we have said previously, the proper approach in designing numerical indicators to assist us in military computation is to first decide which criterion these methods of establishing the worth scales should meet and to then design them accordingly, since the degree of measurement involved in military-worth studies is necessarily limited.[14]

CONCLUSIONS

As a series of concluding remarks, we may ask whether there is any procedure by which we may by-pass the difficulties outlined above and arrive at the final indicator of multi-dimensional priority by some other means than by building them up from a combination of their various components. One possible approach has been suggested by Professor Thomson Whitin [19]. Let us imagine a military game in which certain of our forces are opposed to forces of the enemy. The value of the game to us can be calculated, as well as the additions to this value which results from having additional forces in action. Thus, we have a numerical measure of the value of a particular weapon or unit in a particular position which was directly derived, rather than having been determined by the methods we have mentioned above, and which can be used to guide civilian production, to determine the relative emphasis to be placed on the production of various sorts of weapons, and so on. The difficulty in this sort of approach is that it really only pushes the priority-determination problem back one step, since it then becomes necessary to determine the number within the game matrix. Here, as throughout the general question of indicators of military worth, it is implausible to assume that we shall ever be able to give these indicators truly numerical properties. The

[13] In simplest terms, the principle of additivity seems ill-adapted to this method because it will by definition give, for example, the value of a particular piece of equipment in a particular assignment, or the sum of the value of the equipment and of the place or operation to which it is assigned. Thus, if a very valuable piece of equipment is assigned to a worthless operation, the value of this assignment would be the value of the equipment itself if we were to add; whereas, in fact, the value of this assignment is patently zero. The process of multiplication, on the other hand, introduces (even though imperfectly) the concept that the value of the equipment in a particular operation is a unique alternative and depends on the value of the equipment and of its location in a fairly complex fashion. See [1] for a more extended discussion.

[14] In a more general sense, it can be shown that an adoption of the procedures and problems introduced into economic literature by Professor Kenneth J. Arrow (see [18]) can be used to show that any method of resolving the multidimensional priority problem, in which we have less than absolute-zero scales of measurement, will fail to meet a set of very reasonable criteria, see [1]. Also see in reference to the approach suggested in footnote 8.

proper approach in this instance would seem to be a redirection of game-theory research towards a determination of whether meaningful game theoretical models can be built on a foundation other than that of comparable numerical utilities. It can be easily seen that if one possesses mere orderings of objectives by the competing parties, little can be said about the solution of the game [20]. But a comparison of the sensitivity of the outcome of the game to changes in the matrix of coefficients compared with the degree of measurement which we possess on the ranking of alternatives by the competing parties may yield us important results.

REFERENCES

[1] W. G. Mellon, "An Approach to the General Theory of Priorities," to be published.

[2] J. W. Smith, "A Plan to Allocate and Procure Electronic Sets by the Use of Linear Programming Techniques and Analytical Methods of Assigning Values to Quantitative Factors," Naval Research Logistics Quarterly, Vol. 3. No. 3, Sept. 1956, pp. 151-162.

[3] George Suzuki, "Procurement and Allocation of Naval Electronic Equipments," Naval Research Logistics Quarterly, Vol. 4, No. 1, March 1957, pp. 1-8.

[4] R. J. Aumann and J. B. Kruskal, "The Coefficients in an Allocation Problem," Naval Research Logistics Quarterly, Vol. 5, No. 2, June 1958, pp. 111-123.

[5] _____ "Assigning Quantitative Values to Qualitative Factors in the Naval Electronics Problem," Naval Research Logistics Quarterly, Vol. 6, No. 1, March 1959, pp. 1-6.

[6] Henry Solomon, Joseph P. Fennell, and Marvin Denicoff, "A Method for Determining the Military Worth of Spare Parts," Serial T-82/58, Logistics Research Project, The George Washington University, Washington, D.C., April 1958.

[7] Henry Solomon, "The Determination and Use of Military Worth Measurements for Inventory Systems," these proceedings.

[8] P. L. Folsom, "Military Worth and Systems Development," these proceedings.

[9] Oscar Morgenstern, "Consistency Problems in the Military Supply System," Naval Research Logistics Quarterly, Vol. 1, No. 4, December 1954, pp. 265-281.

[10] Henry E. Eccles, Logistics in the National Defense, Harrisburg, 1959, pp. 153-159.

[11] Sidney Siegel, "A Method for Obtaining an Ordered Metric Scale," Psychometrika, Vol. 21, No. 2, June 1956, pp. 207-217.

[12] R. J. Aumann and J. B. Kruskal, "Assigning Quantitative Value by Qualitative Methods in an Allocation Procurement Problem," Technical Report 21, Analytical Research Group, James Forrestal Research Center, Princeton, N.J., July 1956.

[13] Johann Pfanzagl, "A General Theory of Measurement - Applications to Utility," Research Memorandum No. 5, Econometric Research Program, Princeton University, Princeton, N.J., December 8, 1958.

[14] Gerard Debreu, "Representation of a Preference Ordering by a Numerical Function," in R. M. Thrall, C. H. Combs, and R. L. Davis, eds., Decision Processes, Wiley and Sons, New York, 1954; reprinted as Cowles Commission Paper, New Series, No. 97, 1954.

[15] Melvin Hausner, "Multidimensional Utilities," in Thrall, Combs, and Davis, eds., Decision Processes, Wiley and Sons, New York, 1954.

[16] R. M. Thrall, "Application of Multidimensional Utility Theory," in Thrall, Combs, and Davis, eds., Decision Processes, Wiley and Sons, New York, 1954

[17] Tjalling C. Koopmans and Martin J. Beckman, "Assignment Problems and the Location of Economic Activity," Econometrica, Vol. 25, No. 1, January 1957, pp. 53-76.

[18] Kenneth J. Arrow, Social Choice and Individual Values, Wiley and Sons, New York, 1951.

[19] Thomson M. Whitin, The Theory of Inventory Management, 2nd Edition, Princeton University Press, 1957; Chapt. 9, pp. 189-207.

[20] Lloyd S. Shapley and Martin Shubik, "Solution of N-Person Games with Ordinal Utilities," (abst.), Econometrica, Vol. 21, No. 2, April 1953, p. 348.

* * *

THE DETERMINATION AND USE OF MILITARY-WORTH MEASUREMENTS FOR INVENTORY SYSTEMS*

Henry Solomon

Logistics Research Project

In this paper an attempt is made to emphasize the significance of military worth measurements for inventory problems and to point to such measurements which are presently available. The general results and implications of these measurements are discussed.

A significant feature of the inventory problem in the military establishment is its very close connection with the problem of military worth or strategic value. Inventories are held in order to support some current strategic plan and to permit the feasibility of future and uncertain strategic plans. Recognizing the significance of military worth is not new. The development of inventory theory with reference to military problems has included in a fairly explicit way the need for a concept and measurement of military worth. What remains for the future is the task of clarification of the concept as well as the derivation and application of military-worth measurements. The purpose of this paper is to describe and discuss some beginnings in this area.

There are three important general points worth mentioning at the outset, since these points will pervade the discussion. The first is that while the military-worth problem is complex, this is at least matched by the importance of doing something specific about the problem. If one chooses to ignore it, the result is to implicitly postulate that all items in an inventory system are of equal worth under all situations. This is a condition which prevails in actual inventory systems. A second point is that an accumulation of information with regard to the nature of military worth and how the concept should be handled has an important bearing on the structure of inventory models. As more is learned of military worth and its treatment modified, so must the inventory models themselves be modified. The concept and measurement of "worth" must be consonant with other aspects of the model. The third point is that "worth" measurements are not only necessary as part of a model for determining line-item stock levels, etc., but also for the evaluation of the results of a model. In order to determine whether an inventory policy is acceptable, or to solve the more difficult problem of determining which of several inventory policies is best, a rule or evaluation function must be employed which includes "worth" considerations in a well-defined manner.

In the recent past, some specific attention has been given to military worth in inventory problems. The orientation of most of this effort has been the problem of logistics support for small independent combatant units (e.g., an individual ship or a single aircraft squadron). Two

*Research was supported by the Office of Naval Research.

major studies in this area have been in connection with Rand's Fly-Away Kit Program [1] and ALTP [2]. The principal objective of these programs is to maximize the "worth" of items to be stocked for a given operational period, subject to either a space or a weight constraint. Some features are common to both studies with regard to "worth" considerations. Both set out to discover the relative importance and effect of a shortage of a particular part on the mission of the combatant unit. Also, in both studies a questionnaire approach was employed to obtain this information.

Two major differences between the studies are the level at which the questions were asked and the use made of the answers. The Rand questionnaire asked the effect of a part shortage on the mission of the aircraft. ALTP asked questions on two levels, each represented by a separate questionnaire. One questionnaire was directed at learning the effect of a part shortage on the component in which it is installed. A second questionnaire was used to determine the worth of the component to the mission of the ship. The worth of a spare part, then, is a function of its importance to the parent component and, in turn, the importance of that component to the ship's mission.

The second major difference is not how the information was obtained, but how the results were used. This difference may be explained by noting why these studies were conducted. Rand's study was employed to obtain the data necessary for the Fly-Away Kit Model. The basic feature of this model is to stock items on the basis of expected demand, military worth, and weight. The important point, for this paper, is that military worth is to be represented by some numerical value, in order that it may be multiplied by expected demand. Hence, based on the answers obtained from the questionnaire, each item is assigned an arbitrary numerical value representing the relative importance of the item.

In the case of ALTP, the motivation and orientation of the military-worth study was quite different. Here the need to obtain military-worth information became acute due to the observed nature of demand for spare parts aboard ship. This demand for individual items aboard ship is extremely low and sporadic. Based on the large amounts of usage data which were analyzed prior to the military-worth study, it seemed that the situation was one of "uncertainty," as distinct from "risk" where some probability distribution may be readily assumed. Because of this observed nature of demand, it was concluded that the decision as to the range of items to be stocked should receive greater emphasis than the determination of the depth of each item to be stocked. First consideration should then be given to determining the range of items which are necessary for the operation of a mission. This led directly to the military-worth study. The result of the study was that each item was assigned a military-worth category. These categories were not represented by cardinal values, but remained as a system of ordinal measures. Stocking policies were then derived which made great use of these measures, and were in fact built around this military-worth information. Thus, an attempt was made to construct inventory models based on what was learned concerning the nature and significance of a depletion as well as the nature of demand.

One important result of these studies is a verification of the point that not all items are of equal importance. The ALTP study, which focused attention on submarines, demonstrated that the majority of items are of low worth for a single patrol period. This contrasts sharply with present procedures which assume the equal importance of all items. A second, and probably more important, result is that something specific can be done concerning military worth and that it need not remain as one of the intangibles in military-inventory problems.

As may be expected in this subject area, these studies have raised new problems, or have clarified old problems. First, by the accumulation of information regarding the nature and seriousness of inventory depletions, the formulation of inventory problems may be (and are being) reconstructed. Also, one important problem which has been highlighted is that of evaluating inventory models and procedures. This is a problem which has received very little attention. The primary difficulty here is that the evaluation is highly sensitive to the manner is which "worth" information is treated. For example, if arbitrary numerical values are employed, the evaluation of a model will be strongly dependent on these numerical values. With different numerical values, a particular inventory policy may appear better or worse than another. Also, an ordinal system of "worth" measurements prevents difficulties in the process of model evaluation. For example, there remains the question as to how many "low-worth" shortages are equivalent to a "moderately high-worth" shortage. In brief, certainly not all problems have been solved, but some definite progress is in evidence.

Almost nothing has been accomplished in the use of military-worth information for inventory problems which are directly concerned with "system-stocking," provisioning, etc. Again in the case of system-stocking, all items typically are considered to be of equal worth. Thus, for example, all items in an inventory system are assigned the same safety level of supply. Some improvement may be gained by classifying items into various "worth" categories and considering these as a primary criterion for determining safety levels—which will vary depending on the item "worth."

Another use for the type of "worth" information already obtained is in the echeloning of supply. An example of this use would be in the determination of how many items and units are to be placed aboard ship, to be stocked on afloat supply sources (e.g., a tender) and at shore bases. If an item is assigned a "low worth," this may be interpreted to mean that the ship may operate without loss of effectiveness for some given time period without the item. Hence, items such as these may not be stocked aboard ship but rather placed at some single point of supply support.

In provisioning or in the initial procurement of parts for a new component, the kind of "worth" information obtained by ALTP is extremely valuable. This is due to the explicit information concerning the importance of the component, as distinct from the importance of an individual item or part. In the past, proposals for deferred procurement have been offered. Essentially this has meant the deferral of the initial procurement of expensive parts until enough demand experience has been gained. It appears that if deferred procurement is desired one may use the "worth" of the component as a primary consideration for this procedure. That is, to use the "worth" of the component rather than the dollar value of the part. Previously it was mentioned that the demand for parts is extremely low and sporadic. Although at the present time no empirical investigations have been completed to support this, it appears that, although the nature of demand is as stated, most of the demand which does arise is for a relatively small percentage of components. That is, some components are "item-users" while others—probably the majority—are not "item-users." If this is true, and a new component based on function, etc., can be categorized as such, then this information, together with the "worth" of the component, can be utilized for decisions regarding deferred procurement. For example, if a new component is similar to one which is not an "item-user," and this component is "low worth," the procurement of the parts installed in the component may be deferred. This again may be contrasted with the use of the dollar value of an item as the criterion, even though the dollar value may not be related to the "worth" or usage of the item.

In conclusion, some useful beginnings have been made in this area of military worth which previously has been deemed one of the imponderables in the inventory problem. Although much further work is required, some results have already been obtained which are of value to the development of inventory theory as well as to the manager of an inventory system.

REFERENCES

[1] H.W. Karr, "A Method of Estimating Spare-Part Essentiality," Naval Research Logistics Quarterly, Vol. 5, No. 1, March 1958, pp. 29-42.

[2] Marvin Denicoff, Joseph P. Fennell, and Henry Solomon, "A Method for Determining the Military Worth of Spare Parts," Serial T-82/58, The George Washington University Logistics Research Project, Bureau of Supplies and Accounts Allowance List Test Program (ALTP), April 1958.

* * *

PANEL 5
TRANSPORTATION AND SCHEDULING PROBLEMS

SOME PROBLEMS AND MODELS FOR TIME-PHASED TRANSPORT REQUIREMENTS

Chance Constrained Programs with Normal Deviates and Linear Decision Rules

A. Charnes

Northwestern University

and

W. W. Cooper

Carnegie Institute of Technology

This paper discusses the possibilities of analysis of risk and uncertainty in forward planning in time for implementation of transport requirements by means of "chance-constrained" programming models [1]. The possibilities of using chance-constrained programming models, in place of queuing models, and the solution of such a model for the leasing of tanker fleets, is considered in some detail.

INTRODUCTION

This paper continues the studies which were initiated in [2] and extended in [1]. It focuses on a class of chance-constrained programming problems—e. g., those involving linear decision rules in independent normal variates—and develops their direct equivalence to certain mathematical programming problems which can be dealt with in a wholly constructive manner. The procedure thus developed elides, for this class of problems, the "two-stage factorization" of [1]. Preliminary to the development of extensions of this procedure to other classes of variates (or decision rules), a characterization of distributions amenable to such manipulations is obtained in terms of their characteristic (or moment-generating) functions. As will be seen at the end of this paper, however, an extremely simple and clear-cut way of evaluating alterations in quality or risk factors that enter into managerial decisions can be effected for the class of problems considered here. The stage is therefore set for supplying analytical aids to management by tightening or relaxing those parts of the constraints, while avoiding the twin difficulties that are associated with the more usual ex ante approaches to these problems: via either subjective probability or utility analyses. These ex ante approaches either shift the burden of analysis onto management or require management to make preliminary judgments in artificially simplified situations without being able to assess the manner in which their judgments will affect the problems that are actually of interest.

AN EXAMPLE

An example of the class of problems to be considered is one in which a company (an oil company with a tanker fleet, for instance) wishes to determine its ship-chartering policy in the light of its already available fleet.[1] Suppose that the demands for shipping capacity are uncertain in the sense that they are known, from period to period, only to the extent that they arise from normal distributions whose means and variances have specified values. Suppose, further, that the charters for ships fall into one of two classes: spot (or short term) and five year (or long term), and that these are mutually exclusive and exhaustive categories. The charter rates in each period also have known distributions and these, in turn, depend upon the demands for shipping capacity.[2] It is required that the shipping capacity demands be met to at least the levels specified by preassigned probabilities—which may vary from period to period.

Now let there also be given: (1) a stipulated horizon, or planning interval, and (2) a maximal allowance—stated probabilistically—for the number of ship charters which will be in effect on the terminal date. The problem, then, is to find the sequence of decision rules—linear in the demand variates—which are optimal in the sense that they produce minimum total charter costs while honoring the chance constraints that have now been enumerated.

ANALYTICAL FORMULATION

To phrase the problem analytically, let D_1 be the amount of shipping demand in period i,—a variate which is normally distributed with mean $\overline{D}_i$ and variance σ_i^2. Let L_j and S_j be, respectively, the amount of long-term and spot chartering to be undertaken at the beginning of period j. (The index ranges are $i, j = 1, 2, \ldots, N$, where N fixes the horizon.) The charter rates which prevail in period i are represented by $\ell_i(D_i, \ldots, D_i, \ldots, D_N)$ and $s_i(D_1, \ldots, D_i, \ldots, D_n)$ for the long- and short-term charters. To handle long-term charter expirations, the following convention may be employed: $L_j \equiv 0$ whenever $j \leq 0$.

In the formulation which follows, θ_i will be taken to mean the minimum probability which is specified for honoring a constraint in period i. The problem may then be stated:

$$\min \quad E\left\{\sum_{j=1}^{N} (\ell_j L_j + s_j S_j)\right\}$$

$$(1) \qquad \text{subject to} \quad \Pr\left\{\sum_{j=1}^{5} L_{i-j+1} + S_i \geq D_i\right\} \geq \theta_i, \quad i=1, \ldots, N$$

$$\Pr\left\{\sum_{j=1}^{5} L_{N-j+1} \leq L\right\} \geq \theta_{N+1}$$

The functional is, of course, only an analytical restatement of the fact that expected total costs are to be minimized over the N-period horizon. The first set of constraints refers to the

[1] This problem is part of a larger one suggested to us by M. J. Connor, M. D. Chipilane, and G. W. Gershafski of Esso Research and Engineering Co.

[2] The generalization of these effects to supply as well as demand factors will not be considered here.

requirement that the sum of long-term and spot charters shall be sufficient, to at least probability β_i, to cover the demands, D_i, in each period. The final constraint is a representation of the requirement that unexpired charters shall not exceed a level, L, with probability at least β_{N+1}.[3]

Since at the i^{th} period only $D_1, \ldots, D_{i-1}$ are known, any constructive specification of L_i must involve only these D_i.[4] The most general linear decision rule yielding L_i as a function of these D_i's is therefore precisely

$$L_i = \sum_{k=1}^{i-1} \gamma_{ik} d_k + \gamma_i, \qquad i = 1, 2, \ldots, N \tag{2}$$

where

$$d_k \equiv D_k - \bar{D}_k.$$

The $\bar{D}_k$ are known values and their values and their contributions (as well as the possible contribution of known statistics of future D_k's) can be taken up in the variable γ_i.

APPLICATION WITH NORMAL AND STATISTICALLY INDEPENDENT VARIATES

Utilizing (2), then by direct substitution, the functional in (1)—whose expected value is to be made a minimum—can be written

$$E\left\{\sum_{j=1}^{N} \ell_j \left[\sum_{k=1}^{j-1} \gamma_{jk} d_k = \gamma_j\right] + s_j S_j\right\}$$

$$= \sum_{j=1}^{N} \left\{\sum_{k=1}^{j-1} \left[(\overline{\ell_j d_k})\ \gamma_{jk} + \bar{\ell}_j \gamma_j\right] + \bar{s}_j S_j\right\}, \tag{3.1}$$

where the bar over a variable denotes its expected value. The chance constraints may similarly be written

$$\Pr\left\{\sum_{j=1}^{5} \sum_{k=1}^{i-j} \gamma_{i-j+1,k} d_k + \sum_{j=1}^{5} \gamma_{i-j+1} + S_i \geq d_i + \bar{D}_l\right\} \geq \beta_i$$

$$\Pr\left\{\sum_{j=1}^{5} \sum_{k=1}^{N-j} \gamma_{N-j+1,k} d_k + \sum_{j=1}^{5} \gamma_{N-j+1} \leq L\right\} \geq \beta_{N+1} \tag{3.2}$$

$$\Pr\left\{L_i \geq 0\right\} = \Pr\left\{\sum_{k=1}^{i-1} \gamma_{ik} d_k + \gamma_i \geq 0\right\} \geq \alpha_i$$

[3] Lower as well as upper limits can be stated, if desired, say, for the purpose of conducting policy evaluations on either the prescribed levels or the preassigned probability (or risk) coefficients.

[4] Cf. [1] and [2] for further discussion.

where the last set of constraints represents a translation—or generalization—of the usual nonnegative conditions in order to make them suited for this kind of statistical application.

It is now assumed that the $D_k - \overline{D_k} \equiv d_k$ are statistically independent. As may be easily established, [5]—e.g., via the characteristic or moment-generating functions—any linear combination, y, of independent normal variates, x_j, is itself normally distributed as follows: Suppose $y = \sum_1^M a_j x_j + a_o$ and x_j has mean μ_j and variance σ_j^2. Then y has mean $\mu = a_o + \sum_1^M a_j x_j$ and variance $\sigma^2 = \sum_1^M a_j^2 \sigma_j^2$. In the present case letting,

$$y_i = \sum_{j=1}^{5} \sum_{k=1}^{N-j} \gamma_{i-j+1,\,k} d_k - d_i + \sum_{j=1}^{5} \gamma_{i-j+1} + S_i, \quad i=1,\ldots,N$$

(4.1)

$$y_{N+1} \equiv \sum_{j=1}^{5} \sum_{k=1}^{N-j} \gamma_{N-j+1,\,k} d_k + \sum_{j=1}^{5} \gamma_{N-j+1},$$

the constraints are

$$\Pr\left\{y_i \geq \overline{D}_i\right\} \geq \beta_i$$

(4.2)

$$\Pr\left\{y_{N+1} \leq L\right\} \geq \beta_{N+1}$$

and

$$\Pr\left\{L_i \geq 0\right\} \geq \alpha_i$$

for $i = 1, 2, \ldots, N$.

If a normal variate, y, has mean μ and variance σ^2,

$$\Pr\left\{y \geq D\right\} = \frac{1}{\sqrt{2\pi}} \int_{\frac{D-\mu}{\sigma}}^{\infty} e^{-1/2t^2} dt = 1/2 - F\left(\frac{D-\mu}{\sigma}\right) \tag{5.1}$$

where $F(\beta)$ is the so called function—i.e., "erf (β)" -- given by[6]

$$F(\beta) = \frac{1}{\sqrt{2\pi}} \int_0^{\beta} e^{-1/2t^2} dt. \tag{5.2}$$

[5] See any standard statistics reference, such as [3].
[6] Observe that $F(-\beta) = F(\beta)$.

Similarly,

$$\Pr\left\{y \le L\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{L-\mu}{\sigma}} e^{-1/2t^2}\,dt = 1/2 + F\left(\frac{L-\mu}{\sigma}\right). \tag{6.1}$$

Restated analytically, the constraints are then

$$1/2 - F\left(\frac{\overline{D_i} - Ey_i}{\sigma_{y_i}}\right) \ge \beta_i$$

$$1/2 + F\left(\frac{L - Ey_{N+1}}{\sigma_{y_{N+1}}}\right) \ge \beta_{N+1} \tag{6.2}$$

$$1/2 - F\left(\frac{\gamma_i}{\sigma_{y_{N+1}}}\right) \ge \alpha_i,$$

or,

$$F\left(\frac{Ey_i - \overline{D_i}}{\sigma_{y_i}}\right) \ge \beta_i - \frac{1}{2}$$

$$F\left(\frac{L - Ey_{N+1}}{\sigma_{y_{N+1}}}\right) \ge \beta_{N+1} - \frac{1}{2} \tag{6.3}$$

$$F\left(\frac{\gamma_i}{\sigma_{L_i}}\right) \ge \alpha_i - \frac{1}{2}.$$

Since F is a montone increasing functions, by taking the inverse of F, denoted by F^{-1}, we obtain

$$\frac{Ey_i - \overline{D_i}}{\sigma_{y_i}} \ge F^{-1}\left(\beta_i - \frac{1}{2}\right)$$

$$\frac{L - Ey_{N+1}}{\sigma_{y_{N+1}}} \ge F^{-1}\left(\beta_{N+1} - \frac{1}{2}\right) \tag{7.1}$$

$$\frac{\gamma_i}{\sigma_{L_i}} \ge F^{-1}\left(\alpha_i - \frac{1}{2}\right)$$

or,

$$Ey_i - \overline{D}_i \geq F^{-1}\left(\beta_i - \tfrac{1}{2}\right)\sigma_{y_i}$$

$$L - Ey_{N+1} \geq F^{-1}\left(\beta_{N+1} - \tfrac{1}{2}\right)\sigma_{y_{N+1}} \tag{7.2}$$

$$\gamma_i \geq F^{-1}\left(\alpha_i - \tfrac{1}{2}\right)\sigma_{L_i}.$$

We note, since the σ's are nonnegative, that $Ey_i - \overline{D}_i$ has the same sign as (the known) $F^{-1}\left(\beta_i - \frac{1}{2}\right)$; also γ_i has the same sign as $F^{-1}\left(\alpha_i - \frac{1}{2}\right)$, and $L - Ey_{N+1}$ has the same sign as $F^{-1}\left(\alpha_i - \frac{1}{2}\right)$.

For clarity, the graphs of F and F^{-1} are now shown.

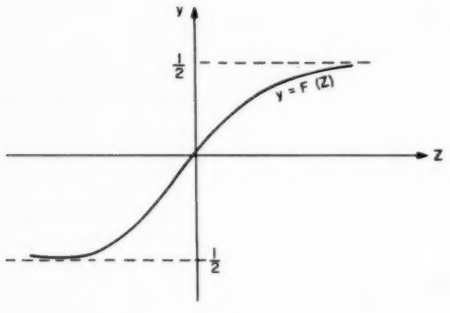


Figure 1 - Graph of y = F(z)

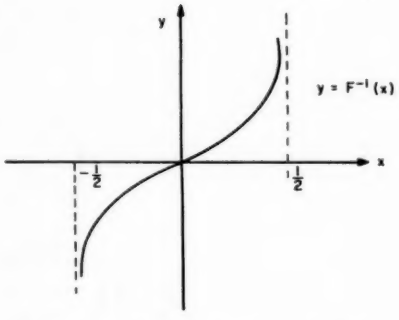


Figure 2 - Graph of $y = F^{-1}(x)$

In a situation such as ours, the β's will usually be large, as is clear from considering the nature of what is involved when a management policy is to be made meaningfully identifiable. In any case, β_i, α_i greater than $\frac{1}{2}$ implies $F^{-1}\left(\beta_i - \frac{1}{2}\right)$, $F^{-1}\left(\alpha_i - \frac{1}{2}\right) > 0$, and thus (a fortiori)

$$Ey_i - \overline{D}_i \geq 0$$

$$L - Ey_{N+1} \geq 0 \tag{8}$$

$$\gamma_i \geq 0.$$

Both sides of (7.2) may thus be squared without altering the status of the inequalities. This gives

$$(Ey_i - \overline{D}_i)^2 \geq [F^{-1}(\theta_i - \tfrac{1}{2})]^2 \sigma^2_{y_i}$$

(9.1)
$$(L - Ey_{N+1})^2 \geq [F^{-1}(\theta_{N+1} - \tfrac{1}{2})]^2 \sigma^2_{N+1}$$

$$\gamma_i^2 \geq [F^{-1}(\alpha_i - \tfrac{1}{2})]^2 \sigma^2_{L_i}.$$

Recalling (4.1), evidently

$$Ey_i = \sum_{j=1}^{5} \gamma_{i-j+1} + S_i$$

(9.2)
$$Ey_{N+1} = \sum_{j=1}^{5} \gamma_{N-j+1}.$$

To obtain $\sigma^2_{y_i}$ and $\sigma^2_{y_{N+1}}$ we must rearrange the expressions for y_i and y_{N+1}, which are of the form

(10.1)
$$y = \sum_{j=1}^{5} \sum_{k=1}^{i-j} a_{i-j+1,k}$$

in terms of k. Thus,

$$y = \sum_{k=1}^{i-1} a_{ik} + \sum_{k=1}^{i-2} a_{i-1,k} + \sum_{k=1}^{i-3} a_{i-2,k} +$$

$$+ \sum_{k=1}^{i-4} a_{i-3,k} + \sum_{k=1}^{i-5} a_{i-4,k},$$

(10.2) or

$$y = \sum_{k=1}^{i-5} \left(\sum_{j=1}^{5} a_{i-j+1,k} \right) + \sum_{j=1}^{4} a_{i-j+1,i-4} +$$

$$+ \sum_{j=1}^{3} a_{i-j+1,i-3} + \sum_{j=1}^{2} a_{i-j+1,i-2} + a_{i-1,i-1}.$$

Thus

$$y = -d_i + \sum_{k=1}^{i-5}\left(\sum_{j=1}^{5} \gamma_{i-j+1,k}\right) d_k + \sum_{j=1}^{4} \gamma_{i-j+1,i-4}\, d_{i-4} +$$

$$+ \sum_{j=1}^{3} \gamma_{i-j+1,i-3} d_{i-3} + \sum_{j=1}^{2} \gamma_{i-j+1,i-2} d_{i-2} + \tag{10.3}$$

$$+ \gamma_{i-1,i-1} d_{i-1} + \text{constants}$$

and

$$y_{N+1} = \sum_{k=1}^{N-5}\left(\sum_{j=1}^{5} \gamma_{N-j+1,k}\right) d_k + \sum_{j=1}^{4} \gamma_{N-j+1,N-4} d_{N-4} +$$

$$+ \sum_{j=1}^{3} \gamma_{N-j+1,N-3} d_{N-3} + \sum_{j=1}^{2} \gamma_{N-j+1,N-2} d_{N-2} + \tag{10.4}$$

$$+ \gamma_{N-1,N-1} d_{N-1} + \text{constants.}$$

Therefore,

$$\sigma_{y_i}^2 = \sigma_i^2 + \sum_{k=1}^{i-5}\left(\sum_{j=1}^{5} \gamma_{i-j+1,k}\right)^2 \sigma_k^2 + \left(\sum_{j=1}^{4} \gamma_{i-j+1,i-4}\right)^2 \sigma_{i-4}^2$$

$$+ \left(\sum_{j=1}^{3} \gamma_{i-j+1,i-3}\right)^2 \sigma_{i-3}^2 + \left(\sum_{j=1}^{2} \gamma_{i-j+1,i-2}\right)^2 \sigma_{i-2}^2 +$$

$$+ (\gamma_{i-1,i-1})^2 \sigma_{i-1}^2$$

while

$$\sigma^2_{y_{N+1}} = \sigma^2_{y_N} - \sigma^2_N$$

(10.5)

$$= \sum_{k=1}^{N-5} \left(\sum_{j=1}^{5} \gamma_{N-j+1,\,k} \right)^2 \sigma_k^2 + \left(\sum_{j=1}^{4} \gamma_{N-j+1,\,N-4} \right)^2 \sigma^2_{N-4} +$$

$$+ \left(\sum_{j=1}^{3} \gamma_{N-j+1,\,N-3} \right)^2 \sigma^2_{N-3} + \left(\sum_{j=1}^{2} \gamma_{N-j+1,\,N-2} \right)^2 \sigma^2_{N-2}$$

$$+ (\gamma_{N-1,\,N-1})^2 \sigma^2_{N-1} ,$$

and also

$$\sigma^2_{L_i} = \sum_{k=1}^{i-1} \gamma^2_{ik}\, \sigma_k^2 .$$

Now let

$$\delta_{ik} \equiv \left(\sum_{j=1}^{5} \gamma_{i-j+1,\,k} \right)^2, \qquad k = 1, \ldots, i-5$$

$$\delta_{i,\,i-4} \equiv \left(\sum_{j=1}^{4} \gamma_{i-j+1,\,i-4} \right)^2$$

(11.1)

$$\delta_{i,\,i-3} \equiv \left(\sum_{j=1}^{3} \gamma_{i-j+1,\,i-3} \right)^2$$

$$\delta_{i,\,i-2} \equiv \left(\sum_{j=1}^{2} \gamma_{i-j+1,\,i-2} \right)^2$$

$$\delta_{i,\,i-1} \equiv (\gamma_{i,\,i-1,\,i-1})^2$$

and

$$\nu_i \equiv (Ey_i - \bar{D}_i)^2 = (S_i + \sum_{j=1}^{5} \gamma_{i-j+1} - \bar{D}_i)^2$$

(11.2)

$$\nu_{N+1} \equiv (L - Ey_{N+1})^2 = (L - \sum_{j=1}^{N} \gamma_{N-j+1})^2$$

from the results previously given in (9.2).

Now from (8) and (9.2),

$$0 \leq S_i + \sum_{j=1}^{5} \gamma_{i-j+1} - \bar{D}_i = \sqrt{\nu_i}$$

(12.1)

$$0 \leq L - \sum_{j=1}^{N} \gamma_{N-j+1} = \sqrt{\nu_{N+1}}.$$

Thus, the constraints are

$$\nu_i - \left[F^{-1}(\beta_i - \tfrac{1}{2})\right]^2 \left[\sigma_i^2 + \sum_{k=1}^{i-1} \delta_{ik}\, \sigma_k^2\right] \geq 0$$

(12.2)

$$\nu_{N+1} - \left[F^{-1}(\beta_{N+1} - \tfrac{1}{2})\right]^2 \left[\sum_{k=1}^{i-1} \delta_{Nk}\, \sigma_k^2\right] \geq 0$$

$$\gamma_i^2 - \left[F^{-1}(\alpha_i - \tfrac{1}{2})\right]^2 \left[\sum_{k=1}^{i-1} \gamma_{ik}^2 \sigma_k^2\right] \geq 0$$

with $\nu_j,\ \delta_{ik} \geq 0$.

It is now assumed that $\overline{\iota_j d_k} \equiv E\, \iota_j d_k = 0$ whenever $j \neq k$. With this assumption the functional (3.1) then becomes

(13)

$$\sum_{i=1}^{N} \bar{\iota}_i \gamma_i + \sum_{i=1}^{N} \bar{s}_i S_i 1.$$

This assumption—viz., $\overline{\iota_j d_k} = 0$, $j \neq k$— therefore admits a considerable simplification. This implies, of course, an absence of statistical correlation between the variates ι_j, the long-term tanker rate in period j, and the deviations in demand $d_k = D_k - \bar{D}_k$ for ship tonnages in periods

$k \neq j$. On the other hand, it need not imply lack of a more general stochastic dependence—e.g., dependence of the ℓ_j on the $\overline{D}_i$, or on other parameters.

Now the γ_{jk} and hence the δ_{ik} are not involved in the functional (13). They may therefore be adjusted in the constraints—without penalty—insofar as this creates advantages for the variables which do appear in the functional. In particular we note that decrease of the $\sum_{k=1}^{i-1} \delta_{ik}\sigma_k^2$ or $\sum_{k=1}^{i-1} \gamma_{ik}^2\sigma_k^2$ can only assist in optimization by increasing the possible range of the ν_i. Thus, in optimizing we can set $\delta_{ik} = 0$, which results in

$$\nu_i \geq \left[F^{-1}\left(\beta_i - \tfrac{1}{2}\right)\right]^2 \sigma_i^2$$

(14.1)

$$\nu_{N+1} \geq 0$$

and also

$$\gamma_i \geq 0.$$

But now taking square roots, the first pair of these last expressions may be replaced by

$$\sqrt{\nu_i} \equiv S_i + \sum_{j=1}^{5} \gamma_{i-j+1} - \overline{D}_i \geq F^{-1}\left(\beta_i - \tfrac{1}{2}\right)\sigma$$

(14.2)

$$\sqrt{\nu_{N+1}} \equiv L - \sum_{j=1}^{5} \gamma_{N-j+1} \geq 0.$$

By this route, the problem becomes one of solving for the "certainty equivalents" (see [2]) in

$$\min \quad \sum_{i=1}^{N} \overline{\ell}_i \gamma_i + \sum_{i=1}^{N} \overline{s}_i S_i$$

subject to

$$\sum_{j=1}^{5} \gamma_{i-j+1} + S_i \geq \overline{D}_i + \sigma_i F^{-1}\left(\beta_i - \tfrac{1}{2}\right)$$

(15)

$$-\sum_{j=1}^{5} \gamma_{N-j+1} \geq -L$$

$$\gamma_i,\ S_i \geq 0,$$

which is seen to be a linear programming problem.

CONCLUSION

It should be observed that (15) is a particularly simple linear programming problem with definite structural features that can be employed advantageously in arriving at solutions. This is, of course, in part a result of some of the simplifying assumptions—e.g., normality and statistical independence—that were used. On the other hand, there is an advantage which accrues from this simple form in the way it lends itself to still further extensions. For instance, by the dual theorem of linear programming [4],

$$(16) \qquad \sum_{i=1}^{N} (\bar{\iota}_1 \gamma_i^* + \bar{s}_i S_i^*) = \sum_{i=1}^{N} \omega_i^* \left[\overline{D}_i + \sigma_i \, F^{-1}\left(\beta_i - \tfrac{1}{2}\right) \right] - \omega_{N+1}^* \, L$$

where the symbol * indicates the values of these variables at an optimum. Note, then, that the dual variables ω_i^*, $i=1,2,\ldots,N$ bear directly on the "levels" $\overline{D}_i$ and the "risks" $\sigma_i F^{-1}\left(\beta_i - \frac{1}{2}\right)$ in each period, while ω_{N+1}^* plays a similar role with respect to the horizon level, L. As is well known, these variables can be interpreted so that, within prescribed ranges, they yield the incremental costs of varying any of the indicated "levels" and "risks." Help is therefore readily available to supply management with systematic analytical guidance. With this help, the managers can then evaluate their "policies" within the context of the actual problem in which their interests lie. In addition, their backgrounds should render them well-suited to make expert and responsible evaluations.

REFERENCES

[1] Charnes, A. and W. W. Cooper, "Chance-Constrained Programming," Management Science 6, no. 1, October, 1959.

[2] Charnes, A., W. W. Cooper and G. H. Symonds, "Cost Horizons and Certainty Equivalents: An Approach to Stochastic Programming of Heating Oil Production," Management Science 4, no. 3, April, 1958.

[3] Cramér, Harald, Mathematical Methods of Statistics, Princeton: Princeton University Press, 1946.

[4] Gale, D., H. W. Kuhn and A. W. Tucker, "Linear Programming and the Theory of Games," in Activity Analysis of Production and Allocation, Cowles Commission for Research in Economics, Monograph No. 13, T. C. Koopmans, ed. New York: John Wiley & Sons, Inc., 1951.

* * *

DETERMINING MERCHANT SHIPPING REQUIREMENTS IN INTEGRATED MILITARY PLANNING*

Ralph B. Hunt and Erling F. Rosholdt

The George Washington University
Logistics Research Project

To establish a perspective, a discussion is made of the over-all planning process involved in the determination of military requirements for the merchant shipping needed to support a prescribed military plan. Three of the principal parts of the planning process, namely: the determination of "pure" pipeline requirements, ship repair loads, and repair replacement ships, under convoy and nonconvoy conditions, are examined. After this, the newly developed models, assumptions, and methodologies for these are dissected and illustrated by means of figures.

Typical characteristics and behavior of both the nonconvoy and convoy-type pipelines and the effect of the repair policies upon them are presented. Additional logistics problems, which were generated in the planning process and which require further research, are outlined.

The logistic problems to be described are concerned with the determination of the military requirements for the merchant shipping necessary to support a prescribed military plan. By using this information, plus comparable data on the merchant shipping requirements of the civilian economy under the same plan, a capability study will be made to see whether available merchant ship resources are adequate to meet the requirements.

In order to establish a perspective, the general problem area of the over-all planning process is outlined in Figure 1. (Parts of the process are already in being, others are being worked on, and others are yet untouched. At this point the term "notional ship" is used to mean a standard reference ship in terms of which actual ships can be expressed. This equivalence will be discussed in more detail subsequently.)

We shall now examine in detail the models, assumptions, and methodologies that have been developed for three principal parts of this flow sheet, namely: the determination of "pure" pipeline requirements, ship repair loads, and repair replacement ships.

The military plan, when it permeates the three services, generates certain lift requirements from each service. The lift-requirement data covers cargo (M/T),[1] POL (bulk)(Bbls), personnel (number), delivery schedule, origin (source), and destination.

Measurement ton is the space available for cargo measured in units of 40 cubic feet to the long ton (2240 pounds).

*Research was supported by the Office of Naval Research.
[1] M/T = Measurement ton.

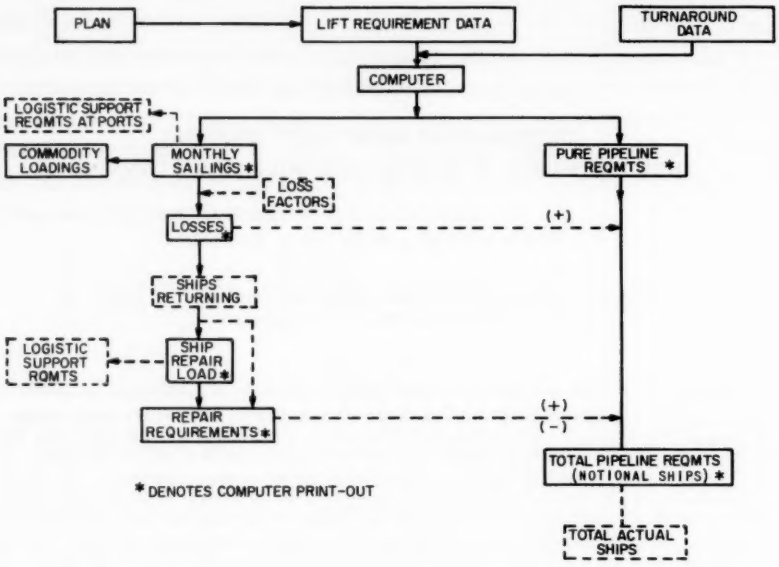


Figure 1 - Determining merchant ship requirements

It is desired to determine how many merchant ships would be needed each time period, under convoy and nonconvoy conditions, to deliver the lift requirements when specified. We refer to these merchant ships as being in a pipeline and define pipeline as follows: The maximum number of ships in use each month to deliver the monthly lift requirements as specified.

Additional data, here called turnaround data (expressed in units of days) are required before the number of ships in the pipeline can be determined. These are:

DISTANCE

CARGO HANDLING
- Loading
- Unloading

BALLAST HANDLING
- Loading
- Unloading

WAITING TIME
- Miscellaneous Delays
- Weather, etc.

CONVOY FACTORS
- Sailing Intervals
- Speed
- Steaming Time
- Schedule.

The necessary elements of a ship's turnaround time are listed in Figure 2.

The determination of the so-called "pure" pipeline requirements is made for two assumed operating conditions: nonconvoy and convoy.

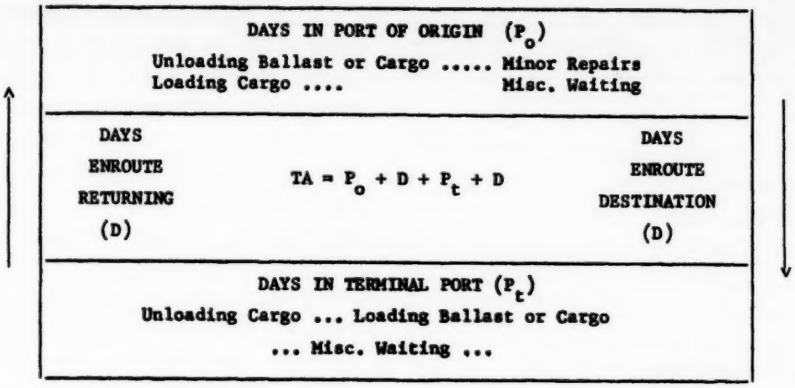

DAYS IN PORT OF ORIGIN (P_o)
Unloading Ballast or Cargo ..... Minor Repairs
Loading Cargo .... Misc. Waiting
DAYS ENROUTE RETURNING (D)
TA = P_o + D + P_t + D
DAYS ENROUTE DESTINATION (D)
DAYS IN TERMINAL PORT (P_t)
Unloading Cargo ... Loading Ballast or Cargo
... Misc. Waiting ...

Figure 2 - Elements of ship turnaround time

Our assumptions for the nonconvoy case are as follows:

1. The pipeline carries only exports from a theater or from the continental United States (CONUS).

2. Lift requirements, expressed in terms of notional ships, are scheduled for 30-day periods beginning with a definite date that coincides with the start of the first time period.

3. Delivery is to be made within each time period if possible, with the day of arrival at the destination being considered the delivery date.

4. No pipeline is in operation over the routes designated by the Lift Requirement Schedule, but ships are available at the port of origin to start the pipeline.

5. Lift requirements are delivered at a uniform average rate by ships sailing daily from the port of origin.

6. Lift requirements level off after some specified number of time periods.

7. Any overdue lift requirement will be combined with the lift requirement current to the particular time period in order to determine the average daily delivery rate.

The first five assumptions need no elaboration. Number 6 makes it possible to reflect accurately those ships, still in the pipeline in the last month of scheduled deliveries, which have not completed their turnarounds. Assumption 7 applies, for example, when the initial leg of the turnaround is greater than 30 days and it is not possible to satisfy Assumption 3.

Graphic models to help us analyze the behavior of a pipeline under these assumptions were developed.

Figure 3 indicates the position of a ship with respect to an assumed turnaround time of 30 days and to the delivery periods. When the ship S, which was available for loading on day 1, arrives at position S_1, it completes the leg (P_o + D) and makes delivery. At position S_2, it has returned home, completing the (P_t + D)-leg and is again available for loading. The cycle is repeated at points S_3 and S_4.

Now if a ship were available for loading on each day of the TA-cycle, it would make delivery (P_o + D) days later. This succession of daily deliveries can be represented as shown in Figure 4.

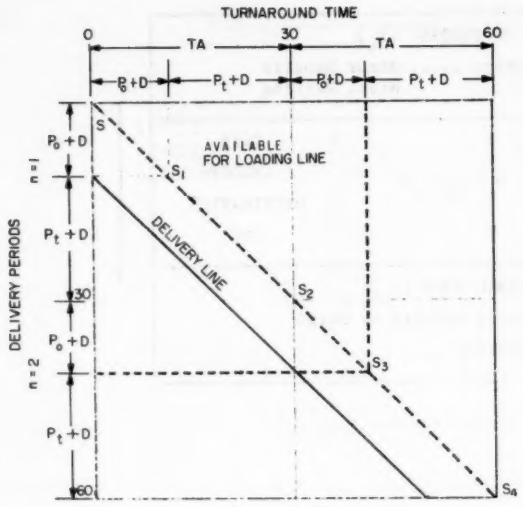


Figure 3 - Position of a ship for a specific turnaround time

In Figure 4, we have taken the second cycle of the TA and placed it under the first cycle. By considering the direction of the movement as downward on the diagram, we now represent our original dashed line as the locus of each ship's availability date for loading, with respect to the TA and the delivery period. The solid line—marked "Delivery Line"—shows the time of delivery of each ship which had started its TA-cycle, $(P_o + D)$-days earlier. The dashed diagonal in the second TA-cycle represents the second time the ships will be back in their port of origin and available for loading, after having completed the return leg of the first TA-cycle. Note that the point on the loading line on the 30th day—the last day of the TA—is the same as that on the zero day for the second cycle, since the first cycle includes the 30th day in the first period.

The preliminary diagraming done in Figures 3 and 4 then led to graphic models similar to the examples shown in Figures 5 and 6.

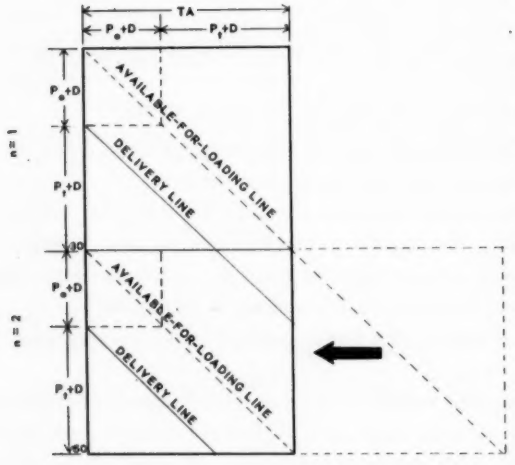


Figure 4

Figure 5 represents both a TA of less than 30 days and three delivery periods of 30 days each. The lift requirement for each period is R_1, R_2, and R_3 notional ships' worth of cargo, respectively. Along the available-for-loading line, each day of the TA is represented by the small circles. The solid, slanting line represents the days on which delivery would be made; i.e., $(P_o + D)$ days after start of TA-cycle. The direction of the movement is downward. Since the (TA = 0) line also represents the last day of the turnaround—as we pointed out in Figure 4—it is less confusing to omit extending the delivery and loading lines to this line after the first TA-cycle is completed. The triangle represents the last delivery that can be made in the first period. By counting along the delivery line, we see that 24 deliveries can be made in the first period. By remembering delivery is inclusive of the 30th day, we see that this is equivalent to 30 - $(P_o + D)$ + 1 or

31 - (P_0 + D) deliveries in the first period. The delivery rate is $R_1/[31 - (P_0 + D)]$ ships per day. The areas marked R_1 on the diagram represent those days in which the ships carrying R_1 requirements are in operation or "in the pipeline." Note that some of the R_1 ships can make more than one delivery in the first period, (9 days worth, specifically).

At the point marked with a square in Figure 5, the first of the 5 groups of R_1 ships that cannot make a double delivery completes its first TA-cycle and is again available for loading. Notice that the next delivery of these ships will be in the second period. The lift requirements are R_2 notional ships, and there are now 30 delivery days possible in this period. Hence, the average delivery rate will be $R_2/30$ ships per day. At the square, then, a decision must be made whether to increase, to decrease or to leave as is, the size of the group of ships available for loading that day. This decision will depend on the relative size of the groups, $(R_1/24)$, $R_1/[31 - (P_0 + D)]$, and $R_2/30$. Notice also, at the left end of the third loading line, that two of the groups of ships that made double deliveries in the first period are again available for loading on the 29th and 30th day. These two groups, together with the other 5 groups, must be adjusted as necessary to make $R_2/30$ deliveries per day. The total pipeline in the first period will then be one of the following, depending on the relative sizes of the delivery rates:

$$Q_1 = \begin{cases} \text{a) } (TA)\dfrac{R_1}{31-(P_0+D)} & \text{if } \dfrac{R_1}{31-(P_0+D)} \geq \dfrac{R_2}{30} \\ \text{b) } (P_t+D)\dfrac{R_1}{31-(P_0+D)} + (P_0+D)\dfrac{R_2}{30} & \text{if } \dfrac{R_2}{30} > \dfrac{R_1}{31-(P_0+D)}. \end{cases}$$

In the second period, notice that ships are in operation with R_1, R_2, and R_3 lift requirements, respectively, on board. The number of ships in the pipeline during this second period will be given by one of the following equations:

$$
Q_2 = \begin{cases}
\text{a) } (P_t+D)\dfrac{R_1}{[31-P_o+D)]} + (P_o+D)\dfrac{R_2}{30} \\
\qquad \text{if } \dfrac{R_1}{[31-(P_o+D)]} \geq \dfrac{R_2}{30},\ \dfrac{R_2}{30} \geq \dfrac{R_3}{30} \\
\text{b) } (TA)\dfrac{R_2}{30} \\
\qquad \text{if } \dfrac{R_2}{30} \geq \dfrac{R_1}{[31-(P_o+D)]},\ \dfrac{R_2}{30} \geq \dfrac{R_3}{30} \\
\text{c) } (P_t+D)\dfrac{R_2}{30} + (P_o+D)\dfrac{R_3}{30} \\
\qquad \text{if } \dfrac{R_2}{30} \geq \dfrac{R_1}{[31-(P_o+D)]},\ \dfrac{R_3}{30} \geq \dfrac{R_2}{30} \\
\text{d) } (P_t+D)\dfrac{R_1}{[31-(P_o+D)]} + (30-2TA)\dfrac{R_2}{30} + (2TA+P_o+D-30)\dfrac{R_3}{30}. \\
\qquad \text{if } \dfrac{R_1}{[31-(P_o+D)]} \geq \dfrac{R_3}{30},\ \dfrac{R_3}{30} \geq \dfrac{R_2}{30} \\
\text{e) } \dfrac{(2TA+P_t+D-30)R_1}{[31-(P_o+D)]} + (30-2TA)\dfrac{R_2}{30} + (P_o+D)\dfrac{R_3}{30} \\
\qquad \text{if } \dfrac{R_3}{30} \geq \dfrac{R_1}{[31-(P_o+D)]},\ \dfrac{R_1}{[31-(P_o+D)]} \geq \dfrac{R_2}{30}.
\end{cases}
$$

When the initial leg of the TA is greater than 30 days and the two legs are unequal, the graphic model will look like Figure 6.

There are only two points to note about Figure 6. First, the R_1 requirements have been combined with the R_2 requirements to give an average combined delivery rate, since the R_1 requirements cannot be delivered in the first period. Second, the number of deliveries possible in the second period is $[61 - (P_o+D)]$, which corresponds to the $[31 - (P_o+D)]$ deliveries possible when the initial leg was equal to, or less than, 30 days.

An interesting aspect of the graphic analysis was the empirical discovery that the entire range of turnarounds, comprising some 3600 possible combinations over the limits of $(P_o+D)_{max}$ = 60 days, $(P_t+D)_{max}$ = 60 days, and $(TA)_{max}$ = 120 days, could be represented by 28 sets of equations. Figure 7 illustrates the sets and their respective boundaries.

This figure shows that, for turnarounds above 15 days, the set boundaries can be expressed in terms of $(TA + P_o + D)$ and $(TA + P_t + D)$ with respect to multiples of the time period; i.e., 30, 60, 90, or 120 days. That portion of the diagram marked A - A and representing the turnarounds below 15 days has been expanded in Figure 8.

In the range of turnarounds shown in Figure 8, the set boundaries take a different form, being in terms of multiples of the turnaround plus (P_o+D) or (P_t+D). It is interesting to note that a set boundary exists at TA's that are factors of the 30-day time period, namely: 5, 10, 15, and 30; and also (as shown in Figure 7) at multiples of 30.

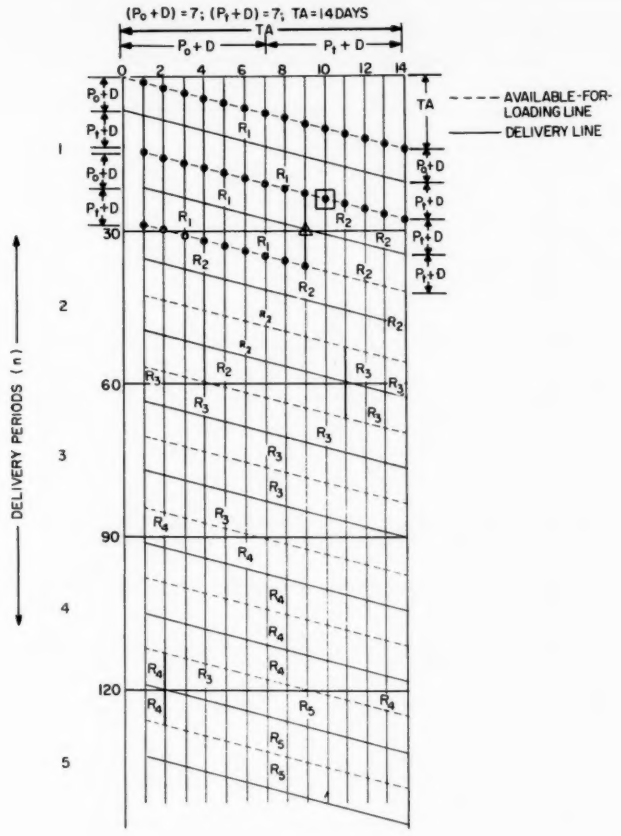


Figure 5 - Graphic Model (TA = 14)

Before discussing the pipeline determination for the convoy condition, it seems desirable to discuss the concepts of notional ships, notional value and—one not mentioned heretofore —notional distance; all of which form a very important part of our methodology. The concepts of notional ship and notional value are not original with us, however, we did originate the concept of notional distance. A notional ship is a reference ship of a fixed capacity, cargo-handling capability, and speed, in terms of which an actual ship can be expressed. The notional value of an actual ship is a measure of how the actual ship compares to the notional ship; i.e., notional value is the ratio:

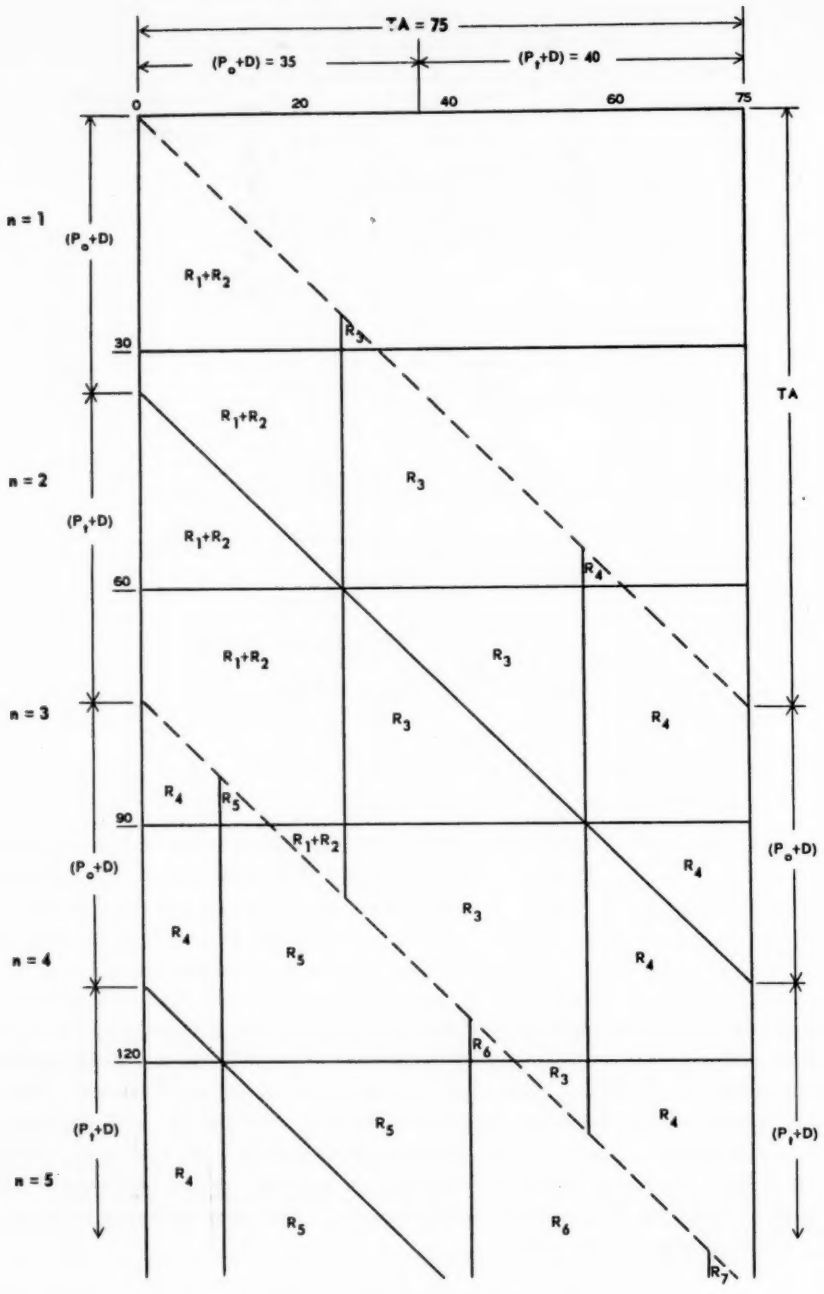


Figure 6 - Graphic Model (TA = 75)

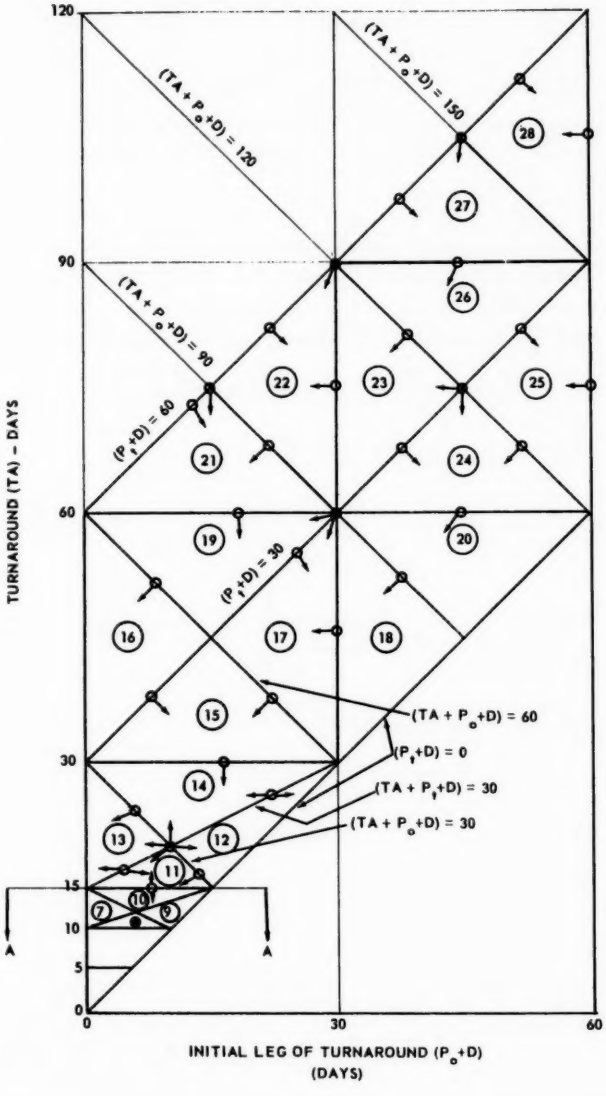


Figure 7 - Boundaries of sets of turnarounds in export-no pipeline system with limits $(P_o + D)_{max}$, $(P_t + D)_{max}$ = 60 days, TA_{max} = 120 days

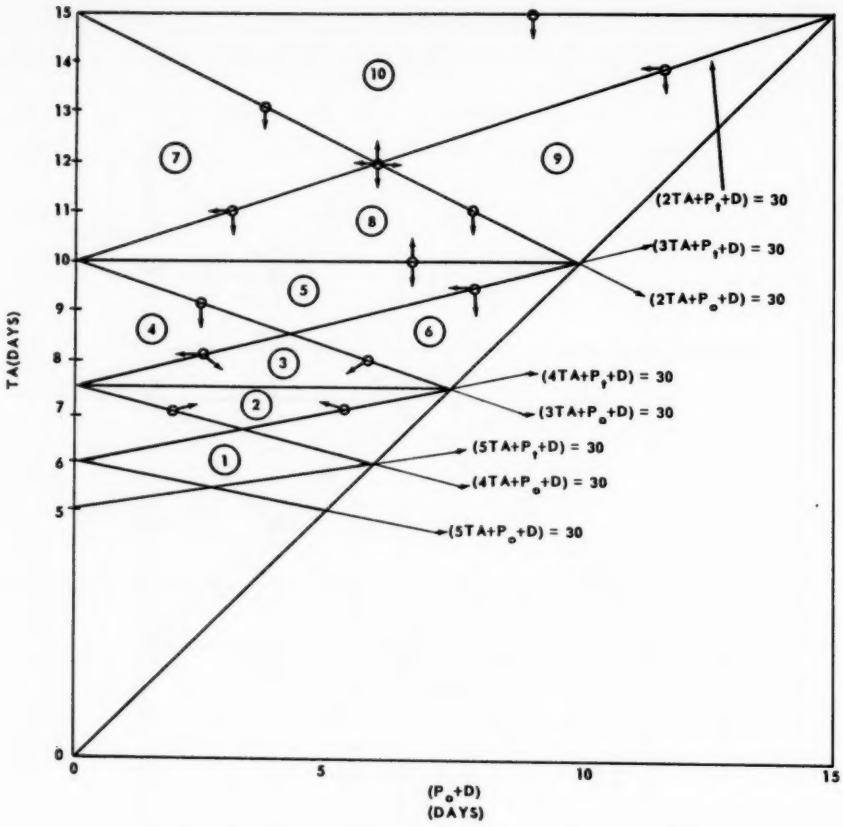


Figure 8 - Expanded view of section A-A from Figure 7

$$NV_a = \frac{C_a}{C_n} \cdot \frac{(TA)_n}{(TA)_a}_r = \frac{C_a}{C_n} \cdot \frac{\left(U_n + L_n + K_n + \frac{D_r}{24S_n}\right)}{(U_a+L_a)\left(\frac{C_a}{C_n}\right) \cdot \left(\frac{H_n}{H_a}\right) + K_a + \frac{D_r}{24S_a}}$$

(Dry cargo or bulk liquid)

$$NV_a = \frac{C_a}{C_n} \cdot \frac{(TA)_n}{(TA)_a}_r = \frac{C_a}{C_n} \cdot \frac{\left(U_n + L_n + K_n + \frac{D_r}{24S_n}\right)}{\left[(U_a+L_a)\frac{C_a}{C_n} + K_a + \frac{D_r}{24S_a}\right]}$$

(Passenger Ships)

The subscripts a and n represent actual and notional ships, respectively. The subscripts $\frac{n}{r}$ and $\frac{a}{r}$ denote, respectively, a notional ship over an actual route and an actual ship over an actual route. The C and TA are capacities and turnaround time, the U and L are unloading and loading times, the D is the two-way distance, and the S is the speed. Therefore, the fraction D/24S is the steaming time in days. H denotes the number of hatches and K is the sum of the waiting times at the origin and destination plus the times for the loading and unloading of ballast, if used. The differences in the two sets of formulas is that the ratio of the hatches is omitted from the lower set. These equations say simply that the notional value of an actual ship will vary directly as the cargo capacities and inversely as the turnaround times, i.e., a greater capacity and lower turnaround time in an actual ship will increase its "value" over the notional ship.

The concept of notional distance is a part of the capability study in which the ship requirements and the ship availabilities are compared. Essentially, the notional distance is a standard reference distance and, since the notional value of an actual ship is a function of the distance (or route) over which it operates, what is done is to convert the notional ship requirements, for the different routes, to the notional ship equivalent at the notional distance selected. This is done by the following equation:

$$\frac{(Q)_{nd}}{(Q)_r} = \frac{(TA)^n_r}{(TA)^n_{nd}} = \frac{P + \frac{D_r}{24S}}{P + \frac{D_{nd}}{24S}}$$

$$(Q)_{nd} = \frac{(24S \cdot P + D_r)(Q)_r}{24S \cdot P + D_{nd}} = \text{Number of ships required for notional distance.}$$

When we use a different distance we change the turnaround time. The ratio of the number of notional ships Q_{nd} required for the notional distance to the number Q_r required for the actual route distance, is the ratio of the turnarounds of the notional ship over the actual route to the TA over the notional distance.

After this conversion is done, the notional value of the actual ships can be determined by the ratio equations for the notional value given previously. The results of a computation involving notional distance are shown in Table 1.

TABLE 1
Examples of Notional Ship Equivalence at a Notional Distance

Actual Two-Way Distance (n. miles)	Notional Ships Req'd at Actual Dist. $(Q)_r$	Turnaround Time for Actual Dist. (days)	Turnaround Time Over Notional Dist. (8000 n. m.) (days)	Equivalent Notional Ships at Notional Distance of 8000 n. m. $(Q)_{nd}$
4000	100	17.1	28.2	60.6
6000	100	22.7	28.2	80.3
9000	100	31.0	28.2	109.9
11000	100	36.6	28.2	129.6
			TOTAL	380.4

Table 1 shows that a total notional value (at 8000 miles) of 380.4 ships must be satisfied by available actual ships with a cumulative notional value (at 8000 miles) which is equal to, or greater than, this sum. If the notional value at the notional distance (8000 miles) of a particular type of actual ship was, for instance 1.3, then the number of actual ships of that type that would be required would be 380.4/1.3 = 293.

We will now consider the determination of the pipeline under convoy conditions. Figure 9 shows a schematic diagram of the turnaround cycle of an escort group.

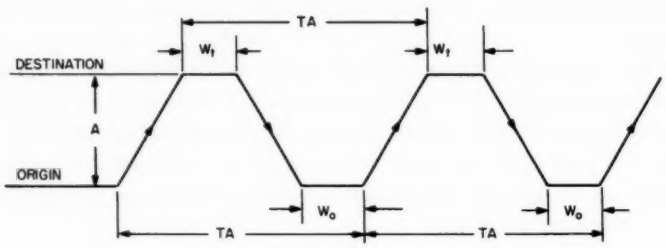


Figure 9 - Turnaround cycle of an escort group

The times in port, W_o and W_t, are short (usually only 3 or 4 days), just long enough for the escorts to refuel, reprovision, handle miscellaneous port delays, and then depart.

Our next graphic model is intended to show how a convoy system must work if the number of escort groups required and the number of merchant ships required in the pipeline are to be minimized. Our studies have indicated that certain relationships between the sailing interval (S.I.), the turnaround times of both escort groups $(TA)_{E.G.}$ and convoy (TA)s, and the time in port of the merchant ships (R_t and R_o) must be maintained if inefficient use of escort groups and unnecessary peaking in the pipeline requirements are to be avoided.

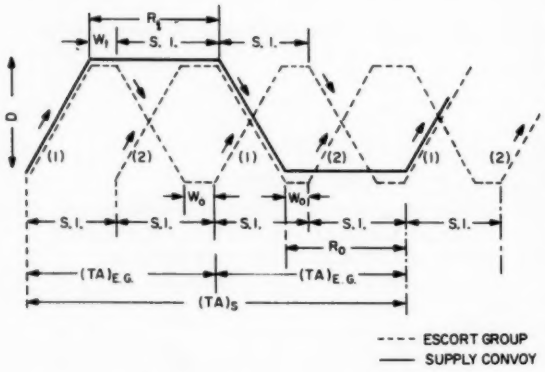


Figure 10 - Convoy scheduling to minimize escort groups

The relationships between the escort-group turnaround and the convoy it escorts are indicated in Figure 10. In this case there are four convoys being escorted by only two escort groups, but for clarity, only one convoy is shown in the diagram. Beginning at the lower left hand corner of the diagram, the cycle shown represents an escort group (Number 1) and a convoy departing. After arriving at the destination, the Number 1 escort group remains for time W_t and returns while the convoy is unloading, refueling, reassembling, and the like. While this was going on, the Number 2 escort group left with a convoy. At the end of its W_t, the Number 2 escort group can return

home, escorting the returning Number 1 convoy and then picking up the Number 3 convoy at the home port, etc.

Notice that both turnarounds must be integer multiples of the same sailing interval (S.I.). The respective port times (R_t and R_o) of the convoy turnaround equal the sailing interval (S.I.) plus the corresponding respective port times (W_t and W_o) of the escort group. When these two relationships hold, we say the turnarounds are "interlocked" and only under these conditions is there a minimum of lost time and excessive increases in the pipeline. When the convoy is ready to sail, an escort group is also ready, and vice versa. Other ramifications of this model have been developed for various types of schedules. These are not immediately germane to our problem, and we will not discuss them here.

The assumptions and the convoy pipeline model that follows incorporate the concept of convoy scheduling which we have just outlined. The assumptions for convoy conditions that are to be used for the convoy pipeline model are:

1. The pipeline carries only exports from a theater or from the continental United States (CONUS).

2. Lift requirements are scheduled monthly beginning with a definite date that coincides with the start of the first time period. Lift requirements are expressed in terms of convoy-notional ships.

3. Deliveries must be made within each time period if possible, with the day of arrival at destination being considered the delivery date.

4. No regular pipeline is in operation over the routes designated by the Lift Requirement Schedule, but ships are available at the port of origin to start the pipeline.

5. The turnaround time (TA) of a ship is the elapsed time, excluding scheduled repair periods, between the starts of successive loading periods from the same port of origin.

6. A cargo ship convoy turnaround time is an integer multiple of the respective convoy sailing interval desired, and the minimum practicable turnaround time for cargo ships over a given route must be adjusted (upwards) accordingly.

7. Tankers and transports will have the same turnaround time as the escort group turnaround time, which must also be an integer multiple of the convoy sailing interval.

8. For odd-number turnarounds, the return leg is taken as one day longer than the outbound leg; for even-number turnarounds the legs are equal.

9. Any overdue lift requirements will be delivered in decreasing order, with the most overdue requirement first, the next most overdue requirement following, etc.

10. Convoys sail at uniform time intervals over a specific route.

11. Lift requirements are divided equally among the convoys making delivery in the particular time period.

In Assumption 2, notice that the lift requirements are expressed in what we call convoy-notional ships. These differ from a standard notional ship only in that the speed assumed is the speed of the convoy rather than a standard ship-speed.

Assumptions 6 and 7 merely point out a condition necessary for interlocking. Assumption 7 also takes advantage of the higher speed capability and faster unloading times of tankers and transports which justify separation of these ships in separate convoys from those with the slower dry-cargo ships. Number 8 is arbitrary but stated to establish a consistent policy. The last assumption (Number 11) is a most important one, as you will see when referring to the pipeline model.

Our analysis, illustrated in Figure 11, shows the kind of decisions that control the adjustments to be made to the basic pipeline.

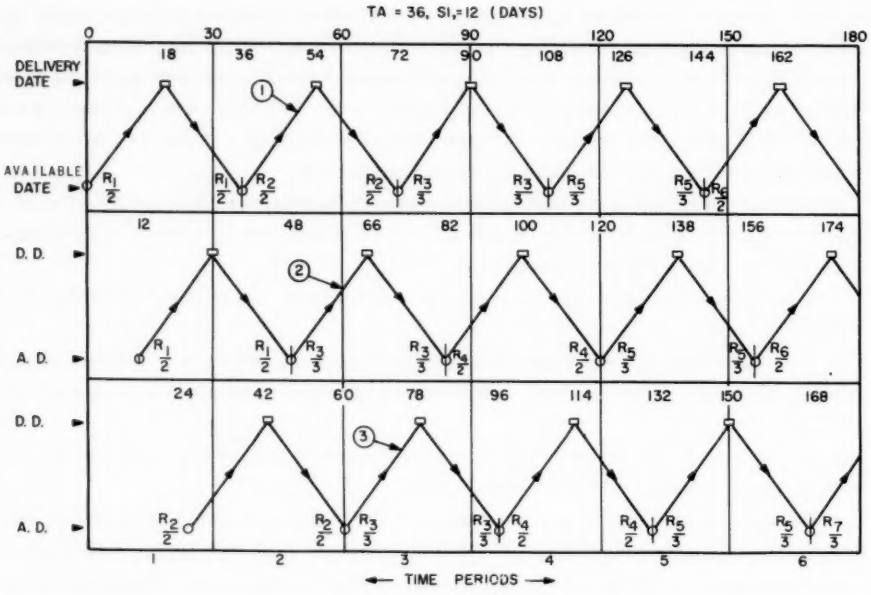


Figure 11 - Convoy pipeline model (3 convoys)

This figure illustrates how the pipeline requirements are derived by making use of the assumption that the lift requirements are distributed uniformly among the convoys arriving at the destination during the delivery period. Notice Convoys 1 and 2 deliver on the 18th and 30th, respectively, and hence the requirements for the first period, R_1, are divided evenly between them. In the second period, Convoy 3 and Convoy 1 (on its second trip) both deliver and hence the R_2 requirements are split between them. Note that the ships for Convoy 3 must be available in the pipeline for the first period on the 24th. In the third period, all three convoys deliver and the R_3 requirements are divided among the three convoys. It can be seen that the decisions to be made at the time that each convoy starts a TA-cycle will depend on comparisons between the number of ships returning and the number needed for the next delivery.

Typical equations for this model are given below. Here the decision rules take the form of inequalities that must be satisfied:

Typical Convoy Pipeline Equations (TA = 36 , S. I. = 12)

Time Period (n)	(Q_n)
(1)	$R_1 + \frac{R_2}{2}$
(2)	$R_1 + \frac{R_2}{2} + \frac{(R_2 - R_1)}{2}$ if $(R_2 > R_1) + \left(\frac{R_3}{3} - \frac{R_1}{2}\right)$ if $\left(\frac{R_3}{3} > \frac{R_1}{2}\right)$
(3)	$\frac{R_2}{2} + \frac{2R_3}{3} + \left(\frac{R_3}{3} - \frac{R_2}{2}\right)$ if $\left(\frac{R_3}{3} > \frac{R_2}{2}\right) + \left(\frac{R_4}{2} - \frac{R_3}{3}\right)$ if $\left(\frac{R_4}{2} > \frac{R_3}{3}\right)$.

Here, Q_n is the number of ships in the pipeline in time period n.

For some combinations of turnaround and sailing interval, it happens that a complete turnaround occurs wholly within a time period. An example of this is shown in Figure 12.

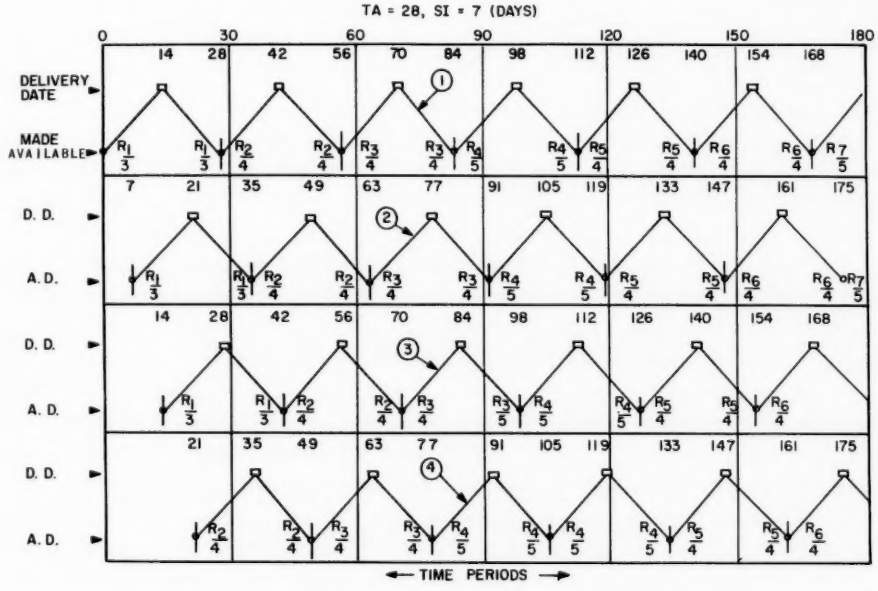


Figure 12 - Convoy pipeline model (4 convoys)

For the Number 2 convoy in the fourth period, a double comparison is required, namely $R_4/5$ vs $R_3/4$, $R_5/4$ vs $R_4/5$, to determine the adjustment to be made.

An interesting discovery from the standpoint of generalizing equations, was that, when the sailing intervals were factors of the 30-day time period (i.e., S.I. = 5, 6, 10, 15 or 30 days), the pipeline formulas could be represented—for a particular combination of convoy turnaround and sailing interval—by a general equation beginning with a time period that was dependent on the turnaround. Table 2 illustrates the form of these generalized pipeline formulas for several turnarounds and a sailing interval of 5 days.

TABLE 2

Typical General Pipeline Equations for
Some Turnarounds at Sailing Interval = 5 Days

Convoy (TA) Days	Time Period (n)	(Q_n)
10	$n \geq 2$	Basic Pipeline + Conditional Adjustment $\overbrace{\frac{R_{n-1}+R_n}{6}}^{\text{Basic Pipeline}} + \overbrace{\left[\left(\frac{R_n-R_{n-1}}{6}\right) \text{ if } (R_n > R_{n-1})\right]}^{\text{Conditional Adjustment}}$
15	$n \geq 3$	$R_{n-1}+\frac{2R_n}{6}+\left[\left(\frac{R_n-R_{n-1}}{6}\right) \text{ if } (R_n > R_{n-1})\right]$
20	$n \geq 3$	$\frac{2(R_{n-1}+R_n)}{6}+\left[\left(\frac{R_n-R_{n-1}}{6}\right) \text{ if } (R_n > R_{n-1})\right]$ $+\left[\left(\frac{R_n-R_{n-1}}{6}\right) \text{ if } (R_n > R_{n-1} \text{ , } R_n \geq R_{n+1})\right]$ $+\left[\left(\frac{R_{n+1}-R_{n-1}}{6}\right) \text{ if } (R_n > R_{n-1} \text{ , } R_{n+1} > R_n)\right]$
25	$n \geq 3$	$\frac{2(R_{n-1})+3(R_n)}{6}+\left[\frac{2(R_n-R_{n-1})}{6} \text{ if } (R_n > R_{n-1})\right]$ $+\left[\frac{2(R_{n+1}-R_n)}{6} \text{ if } (R_{n+1} > R_n)\right]$
30	$n \geq 3$	$\frac{3(R_{n-1}+R_n)}{6}+\left[\frac{3(R_n-R_{n-1})}{6} \text{ if } (R_n > R_{n-1})\right]$ $+\left[\frac{2(R_{n+1}-R_n)}{6} \text{ if } (R_{n+1} > R_n)\right]$
65	$n \geq 4$	$R_{n-1}+R_n+\left[\frac{5(R_{n+1}-R_{n-1})}{6} \text{ if } (R_{n+1} > R_{n-1})\right]$

Consideration is now given to the determination of repair loads, but first, a little background information is necessary. Merchant ships receive two kinds of repair: voyage repairs and overhauls. Voyage repairs are relatively minor repairs that can be done either enroute or while loading or unloading in port. An overhaul is for major repair jobs that normally require use of a drydock. An overhaul generally requires two weeks in peace time and well over two weeks in time of war, if battle damage is incurred.

The procedure that is currently used in computations of the pipeline requirements assumes that in peace time the ship repair time averages 4 percent of a year (15 days) and in war time the ship repair time averages 13 percent of a year (47 days). These percentage figures are then applied against the ships in the pipeline in order to obtain the number of ships that would be going into overhaul during each time period and, hence, have to be replaced in order to keep the pipeline going.

Several things are fundamentally wrong with the method just described. The ships which are furnished to the services to meet their lift requirements will be in operating condition and will not require overhaul until they complete one or more voyages. Ships that are in a pipeline may be loading or unloading in port, or under way, and not necessarily completing a voyage. Hence, taking a percentage of a pipeline inflates substantially not only the number of ships entering overhaul, but also the repair replacements that must be provided to keep the pipeline going. In addition, the current method overlooks the fact that, in the same time period, the ships leaving overhaul could be used to offset the ships entering overhaul, thus reducing the number of repair replacement ships needed.

An alternative repair policy and computation method was developed. From the information gleaned from discussions with knowledgeable Navy people and from available data on shipyard repair times averaged during the last war, a repair policy for planning purposes was proposed.

The basic policy adopted states that under nonconvoy conditions the average repair time per year per ship will be 18 shipdays (5 percent), of which 15 days will be for normal overhaul (similar to peace time) and the balance will be for extraordinary repairs (hazards of navigation, storms, collision, etc.). Under convoy conditions, the average repair time per year will be 47 shipdays (13 percent), of which 30 days is for possible damage, 15 days is for normal overhaul and the balance is for other types of extraordinary damage.

Consideration of the effect of applying this policy to the behavior of nonconvoy and convoy pipeline systems and the time-phasing effects involved led to the establishment of certain assumptions for determining the number of pipeline ships entering overhaul. These assumptions are listed below:[1]

1. After completing a turnaround under the nonconvoy condition, a ship entering overhaul shall require an average of 15 days repair time.

2. After completing a turnaround under the convoy conditon, 2/3 of the ships entering overhaul shall require an average of 15 days repair time, and 1/3 shall require an average of 30 days repair time.

3. The overhaul period begins on TA + 1 day.

4. Ships leaving overhaul shall be placed in an availability pool.

The repair time stated in Assumption 1 is for one overhaul period, whereas, the 18 days repair time cited in the repair policy is per ship per year. Thus, for example, if 18 ships

[1] Assumptions for determining both non-convoy and convoy pipelines are included here.

entered overhaul, all would receive the normal 15 days overhaul. It would also be possible for 3 of these ships to return again within the year for a 15-day emergency overhaul, because of damage due to hazards of the sea. The total repair time per year for those 18 ships would be (18 x 15) + (3 x 15) = 315 shipdays, which is equivalent to 315/18 = 17.5 = 18 days per year per ship.

The graphic models of convoy pipelines in Figures 11 and 12 were used to determine the number of ships that will enter repair during each time period. The emphasis now is in determining how many ships complete their turnaround each period, because it is a fraction of the returning ships that will enter overhaul.

The formula developed for determining the number of notional ships, Q_r, going into repair each half-period (15 days), under nonconvoy conditions, is given below:

$$Q_r = (S) \times \frac{18}{360} \times (TA) \times \frac{1}{15} \times \frac{1}{30} = .000111(TA)(S)$$

where Q_r = notional ships going into repair each half-period,

S = notional ships completing their turnaround in a half-period,

TA = turnaround time in days,

18 = average repair days per ship per year,

360 = days per year (for planning purposes),

15 = repair days for one overhaul period,

30 = least common denominator of fractional 30-day lift requirements comprising (S).

The inclusion of a least common denominator in the formula was an afterthought in order to simplify the tabulation that was made of the formulas.

The formulas for (S) were first determined by an analysis of the nonconvoy pipeline models and then tabulated. A classification of turnarounds into groups was found possible here also. An example, (S) for several sample sets of turnarounds is given in Table 3.

For brevity, the nomenclature used in this table is slightly different. Here, L_o and L_r represent the outgoing and return legs of the turnaround, respectively; corresponding to the $(P_o + D)$ and $(P_t + D)$ we have used up to now. "G" = $R_1/31\text{-}L_o$ corresponding to $R_1/[31\text{-}(P_o+D)]$. The turnaround intervals, to which the equations are applicable, were empirically derived from the models.

TABLE 3

Typical Formulas for Determining (S) - (Non-convoy Condition)

Time Periods \ (TA) Days		1 ≤ TA ≤ 15	16 ≤ TA ≤ 28	29 ≤ TA ≤ 30	31 ≤ TA ≤ 44
1	a	$30(15-TA)G$	-	-	-
	b	$450G$	$30(30-TA)G$	$30(30-TA)G$	-
2	a	$30(L_r+1)G + (14-L_r)R_2$	$30(L_r+1)G + (14-L_r)R_2$	$450G$	$30(45-TA)G$
	b	$15R_2$	$15R_2$	$30G + (29-L_r)R_2$	$30(L_r-14)G + (29-L_r)R_2$
3	a	$(L_r+1)R_2 + (14-L_r)R_3$	$(L_r+1)R_2 + (14-L_r)R_3$	$15R_2$	$15R_2$
	b	$15R_3$	$15R_3$	$15R_2 + (29-L_r)R_3$	$(L_r-14)R_2 + (29-L_r)R_3$
4	a	$(L_r+1)R_3 + (14-L_r)R_4$	$(L_r+1)R_3 + (14-L_r)R_4$	$15\ \ 15R_3$	$15R_3$
	b	$15R_4$	$15R_4$	$15R_3 + (29-L_r)R_4$	$(L_r-14)R_3 + (29-L_r)R_4$
5	a	$(L_r-1)R_4 + (14-L_r)R_5$	$(L_r+1)R_r + (14-L_r)R_5$	$15R_4$	$15R_4$
	b	$15R_5$	$15R_5$	$15R_4 + (29-L_r)R_5$	$(L_r-14)R_4 + (29-L_r)R_5$
6	a	$(L_r+1)R_5 + (14-L_r)R_6$	$(L_r+1)R_5 + (14-L_r)R_6$	$15R_5$	$15R_5$
	b	$15R_6$	$15R_6$	$15R_5 + (29-L_r)R_6$	$(L_r-14)R_5 + (29-L_4)R_6$

NOTE: a and b are the first and second half-periods, respectively.

For convoy conditions, the formula given below for the number of ships, Q_r, entering overhaul in half-periods resembles that for the nonconvoy condition except that the values of the parameters are different.

$$Q_r = (S) \times \frac{47.4}{360} \times TA \times \frac{1}{20} \times \frac{1}{60} = .000108(S)(TA)$$

where Q_r = convoy-notional ships entering repair each half-period,

S = number of ships completing their turnaround each half-period,

47.4 = average repair days per ship per year,

TA = convoy turnaround time of respective ship type (days),

360 = days per year (for planning purposes),

20 = weighted average repair days per overhaul $\left(\frac{2}{3} \times 15 + \frac{1}{3} \times 30\right)$,

60 = least common denominator for fractional 30-day lift requirements comprising S

Formulas for S, as listed by the examples in Table 4, were determined by using the convoy pipeline models and then tabulated.

TABLE 4

Typical Formulas for Determining S - (Convoy Condition)

Time Period		S.I. = 6 Days TA = 36	S.I. = 7 Days TA = 28	S.I. = 10 Days TA = 60
1	a	-	-	-
	b	-	$20R_1$	-
2	a	$40R_1$	$40R_1$	-
	b	$20R_1+12R_2$	$30R_2$	-
3	a	$36R_2$	$30R_2$	$60R_1+20R_2$
	b	$12R_2+12R_3$	$30R_3$	$20R_2$
4	a	$36R_3$	$30R_3$	$20(R_2+R_3)$
	b	$12R_3+12R_4$	$24R_4$	$20R_3$
5	a	$36R_4$	$36R_4$	$20(R_3+R_4)$
	b	$12R_4+12R_5$	$30R_5$	$20R_4$
6	a	$36R_5$	$30R_5$	$20(R_4+R_5)$
	b	$12R_5+12R_6$	$30R_6$	$20R_5$

The fact that we are able to estimate the number of ships entering overhaul each half-period from the pipeline leads us naturally to the problem of determining how many ships must be provided to replace those entering overhaul.

The formula developed for determining the number of repair-replacement ships required under nonconvoy conditions is:

$$Q_{rr} = Q_{r\binom{n}{a}} + Q_{r\binom{n}{b}} - \left[Q_{r\binom{n-1}{b}} + Q_{r\binom{n}{a}} \right] = Q_{r\binom{n}{b}} - Q_{r\binom{n}{a}}$$

where Q_{rr} = net number of repair replacement ships required,

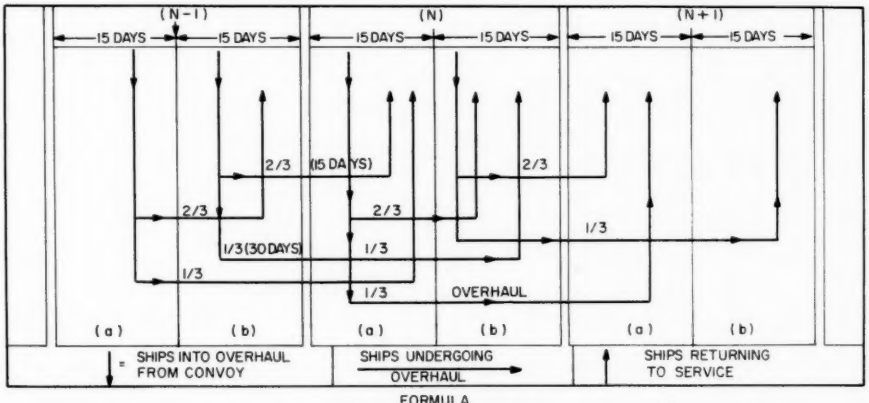
Q_r = number of ships entering overhaul each half-period,

$\binom{n}{a}, \binom{n}{b}$ = first and second half-period, respectively, of the nth time period,

$\binom{n-1}{b}$ = second half-period of the (n-1)th time period.

All this formula does is to provide a balance (based on the 15-day overhaul period) of incoming and outgoing ships each half-period.

For convoy conditions, the determination of the number of repair replacement ships required is more involved. You will recall that the assumption applicable to this case states that 2/3 of the ships entering overhaul remain 15 days and 1/3 remain 30 days. Figure 13 illustrates the model and formula for this operation.



FORMULA

$$Q_{rr(n)} = \left[Q_{r\binom{n}{b}} + 1/3 \; Q_{\binom{n}{a}} - \left(1/3 \; Q_{r\binom{n-1}{a}} + Q_{r\binom{n-1}{b}} \right) \right]$$

Figure 13 - Ship repair replacement (Q_{rr}) under convoy conditions

Some of the characteristics of the pipeline systems we have been discussing are important to understand for two main reasons. One is the implication of what a pipeline means in its demand on resources and military planning. The second is the need for additional research which has been revealed.

Typical characteristics and behavior of the nonconvoy and convoy-type pipelines and the effect of the repair policies upon some of them are shown in Figures 14 through 24. We will comment on only a few of them.

Figure 15 shows how, for three different routes with a combined pipeline requirement, the number of ships required each month as repair replacement ships varies and affects the total pipeline.

Figure 17 illustrates the effect on the pipeline of the repair policy based on percent of pipeline and the proposed repair policy.

In Figures 18 and 19, note that we increased the 24-day cycle to 27 days so that we could obtain a multiple of the S.I. = 9 and the 54-day cycle to 60 days to obtain a multiple of S.I. = 12 days.

By returning to Figure 1, you will see that we have covered only three principal parts of the general problem area represented by the flow diagram shown. The upper left-hand portion requires considerably more work. Even the pipeline section we have discussed has opened up some new avenues for further study.

Just recently we have found indications that, under certain conditions of turnaround and sailing intervals, the size of a pipeline can be greatly reduced, in comparison to that required when sailings are made daily. The development of optimal convoy schedules, although in a classified field, holds considerable promise for logistic improvements. Maximum use of logistic resources in the field of integrated military planning requires a coordination and consolidation of the various methodologies used by the armed services, in order that maximum benefits may be obtained. Empirical methods of logistic research will play an important part in this effort.

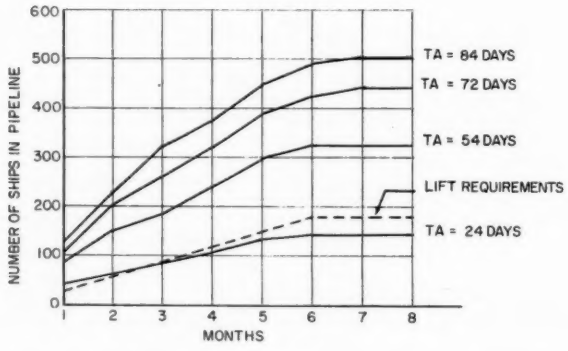


Figure 14 - Effect of pipeline buildup – non-convoy

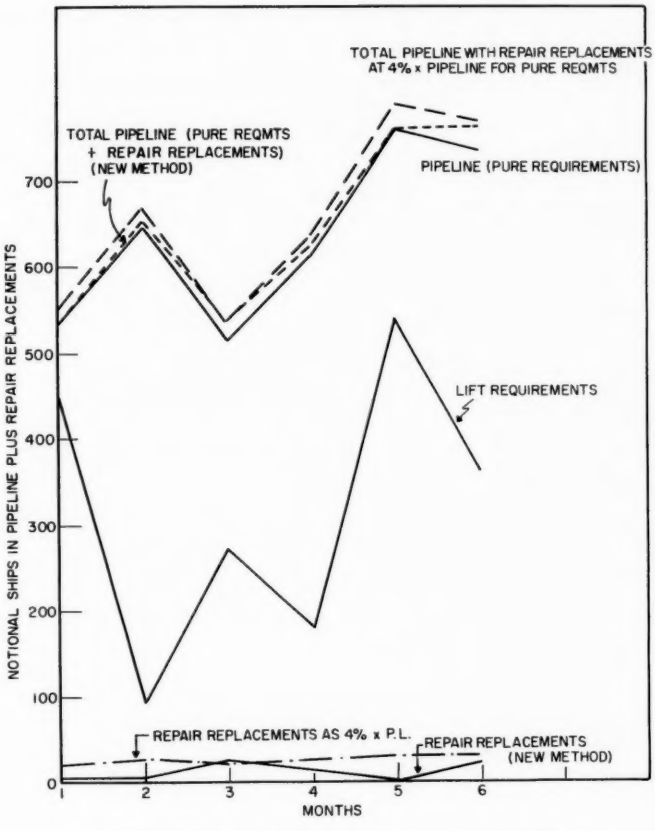


Figure 15 - Effect of sailing intervals on total pipeline buildup with increasing life requirements (using 3 routes with TA of 24, 54, and 72 days

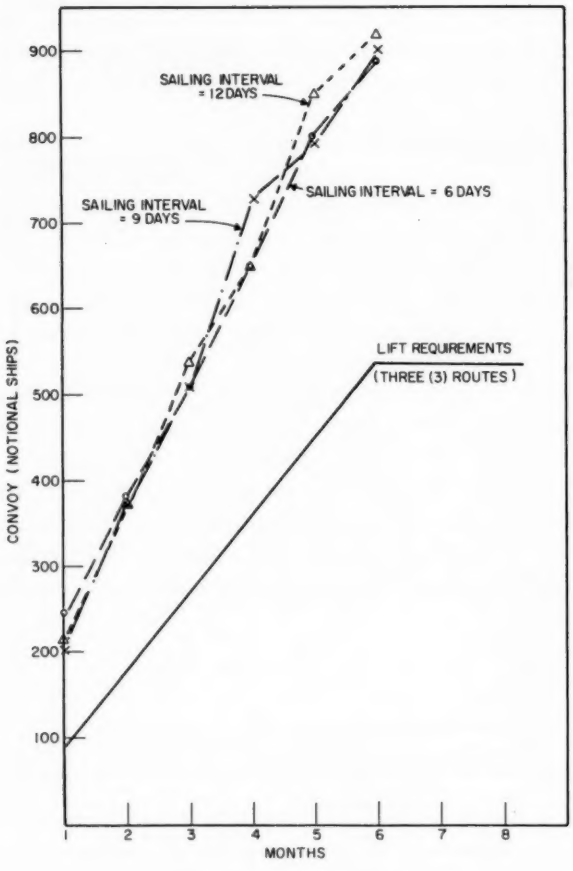


Figure 16 - Effect of S.I. on pipeline buildup - (convoy)

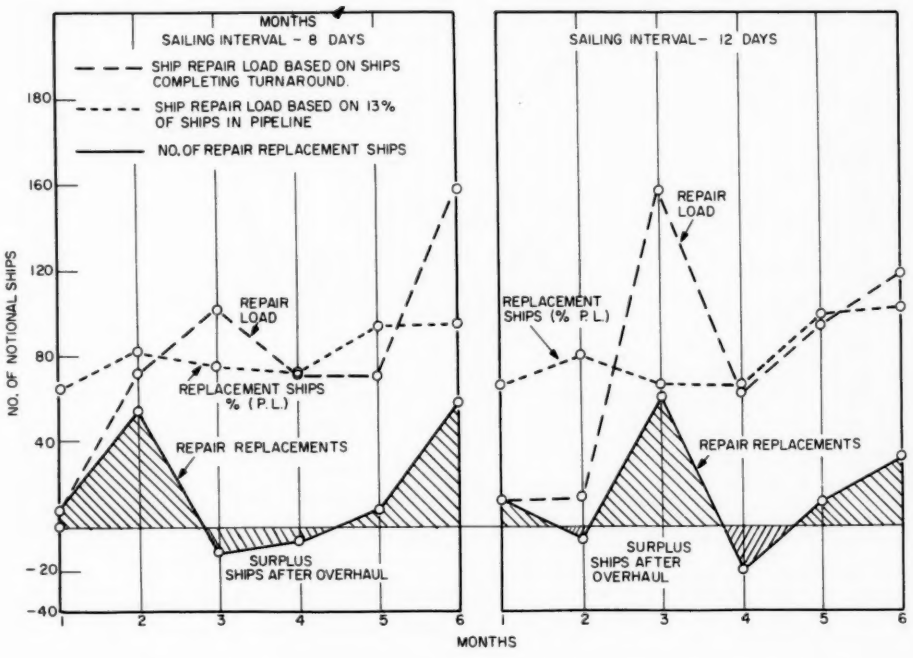


Figure 17 - Combined pipeline requirements with and without repair replacements using variable life requirements under nonconvoy conditions. (TA = 24, 54, and 72 days).

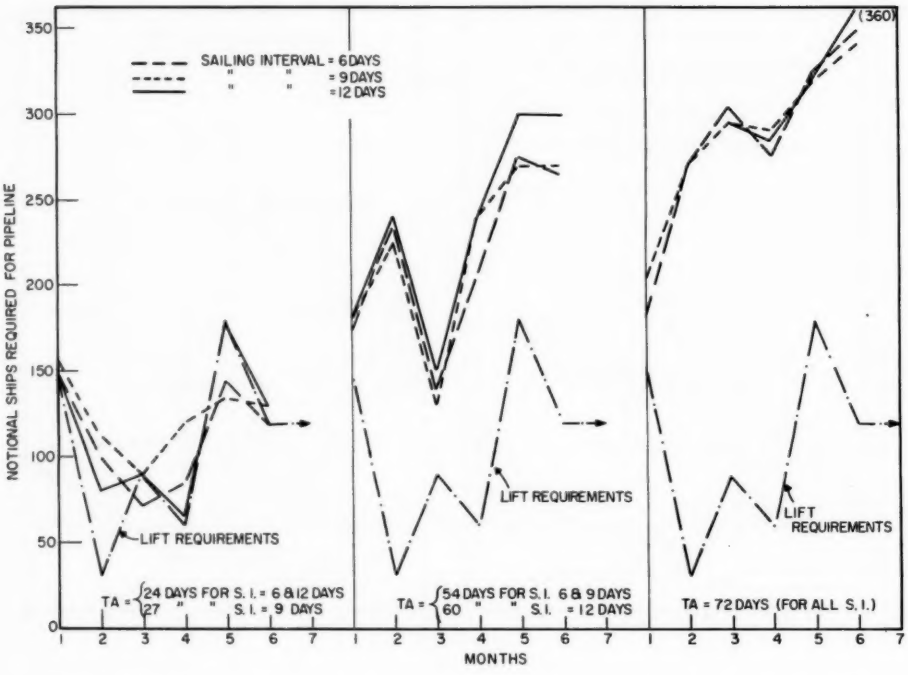

SAILING INTERVAL = 6 DAYS
" " = 9 DAYS
" " = 12 DAYS
NOTIONAL SHIPS REQUIRED FOR PIPELINE
LIFT REQUIREMENTS
TA = 24 DAYS FOR S. I. = 6 & 12 DAYS
27 " " S. I. = 9 DAYS
LIFT REQUIREMENTS
TA = 54 DAYS FOR S. I. 6 & 9 DAYS
60 " " S. I. = 12 DAYS
(360)
LIFT REQUIREMENTS
TA = 72 DAYS (FOR ALL S. I.)
MONTHS

Figure 18 - Effect of convoy turnaround times on ship pipelines (increasing lift requirements and fixed sailing intervals)

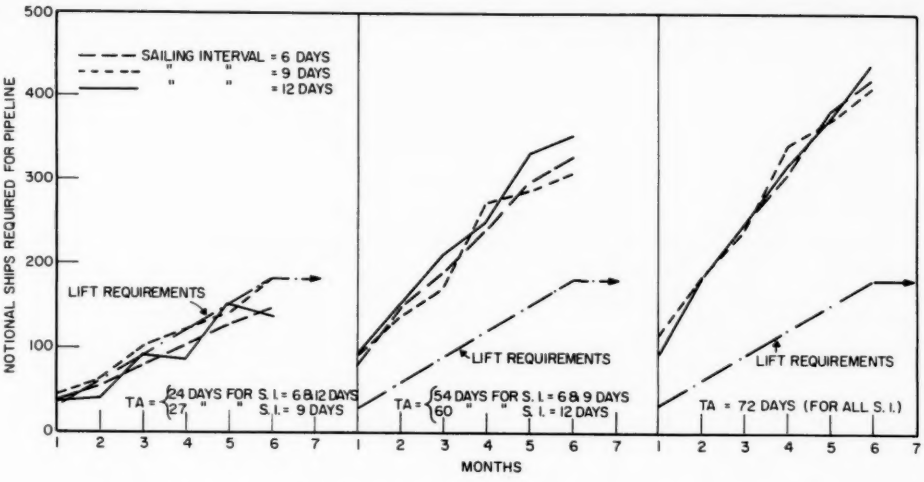

NOTIONAL SHIPS REQUIRED FOR PIPELINE
SAILING INTERVAL = 6 DAYS
" " = 9 DAYS
" " = 12 DAYS
LIFT REQUIREMENTS
TA = 24 DAYS FOR S. I. = 6 & 12 DAYS
27 " " S. I. = 9 DAYS
LIFT REQUIREMENTS
TA = 54 DAYS FOR S. I. = 6 & 9 DAYS
60 " " S. I. = 12 DAYS
LIFT REQUIREMENTS
TA = 72 DAYS (FOR ALL S. I.)
MONTHS

Figure 19 - Effect of sailing intervals on ship pipelines (increasing lift requirements and fixed convoy turnaround times)

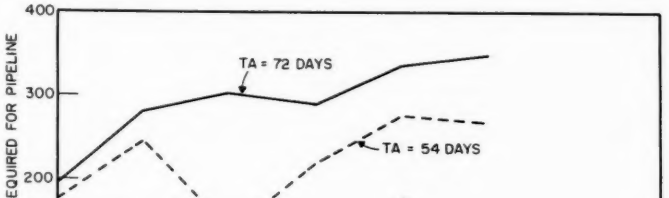


Figure 20 - Effect of TA and variable lift requirements on pipelines (non-convoy)

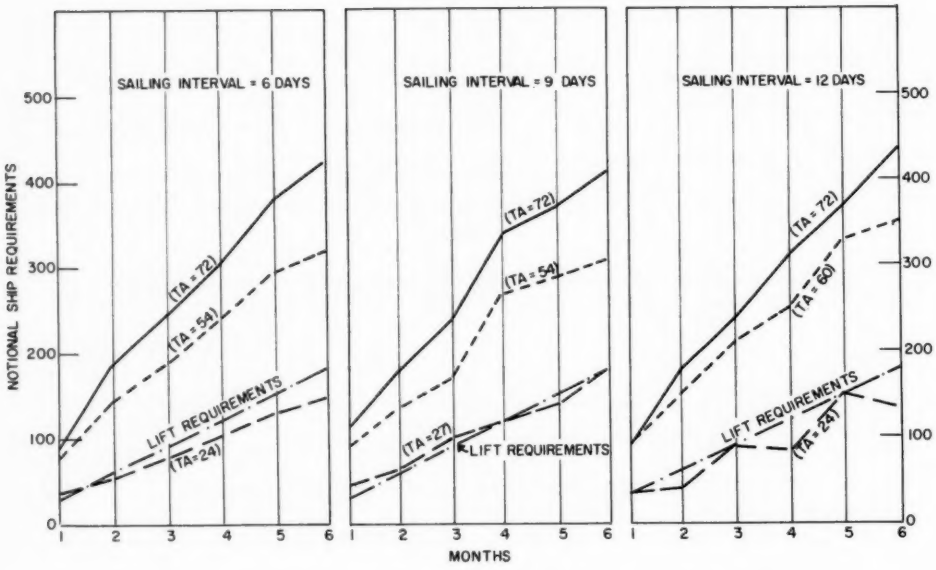


Figure 21 - Effect on pipeline of TA as multiple of S.I.

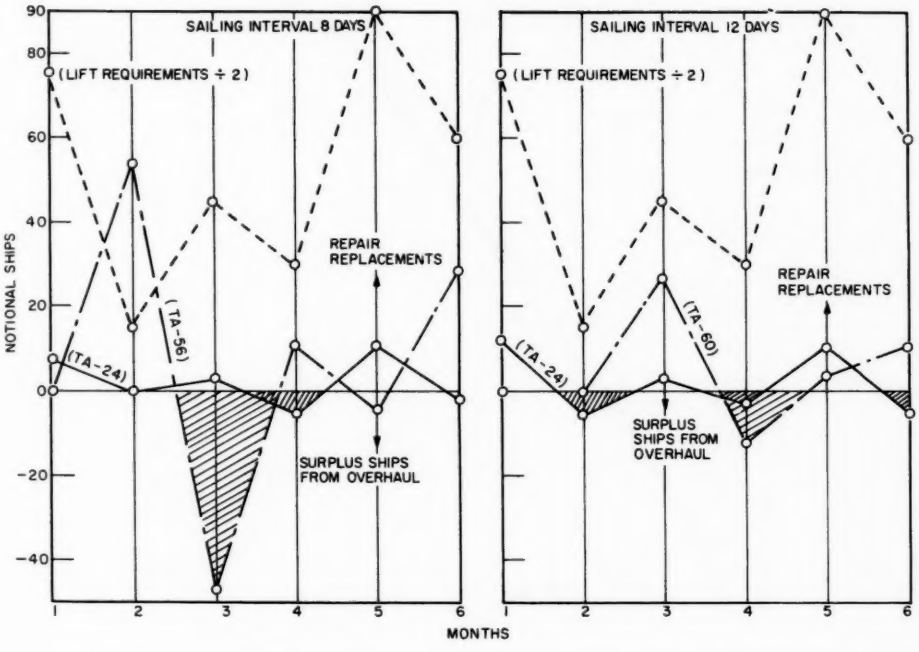


Figure 22 - Effect of convoy sailing intervals on combined ship repair loads and repair replacements (variable lift requirements)

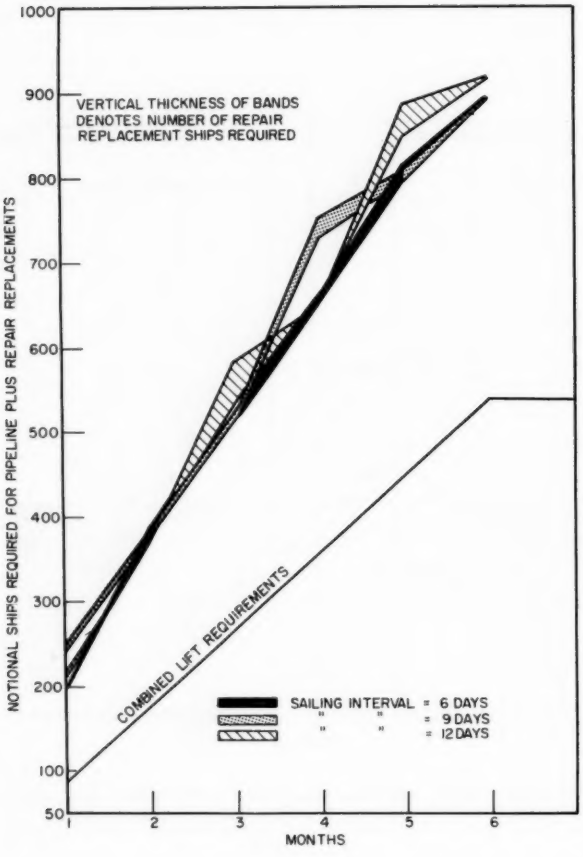


Figure 23 - Effect of sailing interval on total pipeline and repair replacement ships needed over three routes (TA = 24, 54, and 72 days) under convoy conditions (increasing lift requirements)

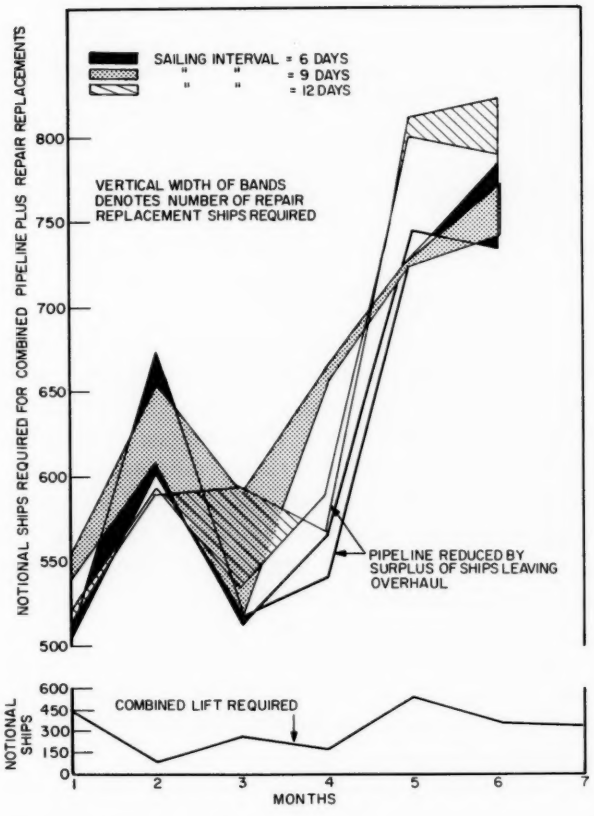


Figure 24 - Effect of sailing interval on total pipeline plus repair replacements needed over three routes (TA = 24, 54, and 72 days) under convoy conditions (variable lift requirements)

SCHEDULING OF CARGO CONTAINERS*

R. R. O'Neill

University of California

Containerization of both military and commercial cargo is being undertaken on a large scale. However, many problems have yet to be solved. An important one is the determination of optimum assignments of containers where less than 100 percent containerization of the ship's cargo must be considered. Mathematical models have been developed which relate possible economic advantages with the number of containers and their allocation to ports and commodities.

INTRODUCTION

Containerization of maritime cargo is not a new concept. Many years ago it was recognized that the heterogeneous nature of dry mixed cargo was a serious impediment to the development of efficient materials-handling systems for the maritime industry. In general, a materials-handling system which incorporates sufficient flexibility to cope with a wide range of sizes, shapes, and weights does so at the expense of efficiency. In particular this is true for the cargo-handling system which involves not only a variety of cargo, but also many handlings on piers and ships which are often designed to other specifications. Some of these specifications are in direct conflict with good materials-handling practices. For example, the requirement of "water tightness" makes the naval architect reluctant to provide large holes in the sides of ships in order to allow a straight-line movement of cargo.

Even though the concept of containerization is not new, the topic is timely because it is only recently that containerization has been undertaken on a large scale. It is being hailed as the biggest thing to hit the waterfront in 20 years. Containerization promises an increased rate of handling with consequent lower handling costs and greater earnings resulting from a decreased turnaround time. It should also be possible to reduce significantly the paperwork which is now required. Containers guard against pilferage and afford protection against damage caused by handling or weather. They are sufficiently sturdy to be used as temporary storage.

Of course, there are disadvantages as well. The principal ones are the cost of purchasing and maintaining the containers; the cost of filling and emptying the containers; and the cost of returning the empties, when necessary. Other disadvantages are the tare weight of the container and the loss of space within and around the container resulting from imperfect fit. Also, special handling equipment may be required. Finally, present tariffs and customs regulations pose special problems. No attempt has been made to enumerate all of the advantages and disadvantages because they are not essential to the discussion of scheduling. They have been introduced in order to point out that the advantages could be overwhelmed by

*This work was supported by the Office of Naval Research under Task NR 347-002.

the disadvantages if the containers are not used properly; and, in this connection, the labor-management relationships which prevail in the ports cannot be ignored. Proper sizing, sound construction, good scheduling, effective controls, and efficient handling will make the difference between profit and loss or success and failure.

Containerization poses many new problems for the maritime industries. Three of these have been studied by the staff of the Cargo Handling Research Project of the Department of Engineering at UCLA. They are: (1) how large should the containers be, (2) how many should be used, and (3) how should they be allocated in order to obtain the greatest "profit" from the use of containers on any particular trade route. It is important to note that the trade route that has been considered is not 100 percent containerized. If every ship carried containers and only containers of a fixed size, the importance of the various parameters which have been explored by the Cargo Handling Research Project would be quite different.

THE DETERMINATION OF OPTIMUM SIZE

Vessels have two capacities: space and weight. If possible a ship is loaded so that it is "full and down" because shipping charges are based upon either the volume or the weight of the shipment, depending upon which is greater. Thus, the density of the cargo is an important factor in the determination of optimum size. Other factors are: the shape, weight, and volume of the commodity unit; the size of the shipment; the fit of the package into the container; the fit of the container into the truck, the railroad car, and the ship; and the capacity of the ship's cargo-handling gear.

By taking these into account, Professor J. D. Carrabino [1, 2] developed a methodology for the determination of an optimum size for shipping containers. He approached the problem on a commodity-by-commodity basis by using a large amount of historical data which had been laboriously compiled from shipping records for eight voyages. The origin, destination, revenue, weight tons, measurement tons, and stowage factor were determined for 1366 commodities. Only 63 percent of these commodities (accounting for 48 percent of the revenue) were containerizable, in the sense that they could be stowed in containers which were no larger than a 10-foot cube. Note that even with the introduction of containers, this trade route would remain less than one hundred percent containerized.

Package dimensions are essential inputs to calculations of fit. Unfortunately these dimensions are not kept as a matter of routine and must be tracked down. Thus, package dimensions were obtained for only a sample of the containerizable cargo, and 54 packages accounted for 64 percent of the revenue from the containerizable cargo.

The percentage of lost inside cube was calculated for each of the 54 package sizes loaded into the 49 different container sizes under consideration. The calculations were carried out on SWAC, the high-speed digital computer at the Numerical Analysis Research Project on the Los Angeles Campus of the University of California. A large number of possible loading patterns were tried. Even when there is no mix of different package sizes within a container, there are many combinations to be considered. The first package may have any one of six initial orientations and there are many different ways of topping off. One hundred and sixty-two different loading patterns were tried for each combination, except when the particular commodity had to be stowed "this side up."

By use of the loading pattern which achieves the least lost cube, the results were tabulated as indicated in Table 1. Note the range from 0 to 42 and the variability. Although

there is a trend toward less lost space as a particular package is loaded into larger and larger containers, the trend is overshadowed by the variation from one container to the next larger one.

TABLE 1

Internal Cubic Loss (in percent) for Typical Commodities and Containers*

Container Size (in.) H D W	Package Size (in.) 10 x 10 x 10	8 x 13 x 19	23 x 11 x 16	24 x 18 x 23
48 x 48 x 48	42.1	31.1	12.2	28.1
60 x 60 x 48	16.7	20.0	8.6	19.5
72 x 72 x 48	21.2	14.2	7.3	4.2
66 x 66 x 66	24.9	12.0	12.7	30.9
78 x 78 x 60	19.5	12.3	5.7	15.6
78 x 90 x 66	18.4	11.7	7.4	12.1
84 x 84 x 78	18.6	6.5	10.6	16.4
84 x 96 x 84	15.0	6.7	9.2	12.0
96 x 96 x 90	12.1	5.7	6.3	4.2
108 x 120 x 96	13.2	7.1	6.3	4.2
120 x 120 x 120	0	7.4	10.3	10.9

*See ref. [1], pp. 40-64.

To determine the cubic loss around containers stowed in irregularly shaped cargo holds, templets of the containers were hand fitted on scaled layouts of the cargo space. It was found that the cubic loss was especially sensitive to the height between decks. Offsetting the lost space and lost revenue is the increased speed of handling which is possible. By assuming that a 6000-pound loaded container could be stowed in the same time as 2000 pounds of loose cargo, and by neglecting the consequences of changes in the amount of pilferage, damage, etc., the savings and loss in dollars were estimated for each commodity. The values calculated for several typical commodities and one container size are shown in Table 2.

THE DETERMINATION OF OPTIMUM SCHEDULES

By use of the concept of the savings index for each commodity, Mr. F. Cesar Toscano [3] developed a procedure for determining how many containers to use and how to allocate these containers among a given group of shipments, in order to maximize the "savings" or "incremental profit" to be realized from handling containers rather than loose cargo. Admittedly, savings indices such as those shown in Table 2 are not readily available. However, they can be calculated and if the calculations are considered to be too tedious, useful estimates can be made. In fact, Professor Carrabino found that experienced Operations Department personnel were able to rank the commodities in terms of decreasing indices rather well. A scheduling procedure for a trade route consisting of N ports was developed by Mr. Toscano in which the cargo could be shipped from any port to any other port, and the ships would always call at the

ports in the same sequence. The cargo which is shipped consists of c different commodities which are containerizable. Furthermore, the type and amount of cargo available at each port is the same for every voyage, so that the total number of containers inbound at each port equals the number outbound.

TABLE 2

Savings Indices for Typical Commodities
Loaded into 6 x 7 x 5-ft Containers*

Commodity	Type of Package	Dollar Savings Per M/T† Over Palletization
Canned pineapple	Wooden case	- 12
Prunes	Wooden case	11
Bottled beer	Carton	4
Canned fish	Wooden case	- 10
Soap	Wooden case	9
Nails and spikes	Keg	- 18
U. S. mail	Bag	7
Tires	Piece	8
Bamboo baskets	Bundle	- 1

*See Ref. 1, pp. 40-64.
†M/T = Measurement-ton.

A linear programming model of this system was formulated and an algorithm was developed to determine the number of containers to be used and the allocation of shipments to containers. The algorithm was then applied to the trade route, consisting of nine ports of call, for which Professor Carrabino had collected the pertinent data.

Mr. Lawrence E. Tannas [4] extended the system studied by Mr. Toscano to include seasonal variations in the amount and type of cargo available for each voyage. In effect, several voyages were lumped together as though they were one. The number of inbound containers at each port must equal the number of outbound containers for the lumped voyages, but not necessarily for each individual voyage. This amounts to a provision for the temporary storage of empty containers at the ports.

The model may be defined as follows:

let a_{ij}^{c} = amount of commodity c available for shipment from port i to port j,

p_{ij}^{c} = savings index for commodity c shipped from port i to port j,

p_{ij}^{e} = cost of shipping an empty container,

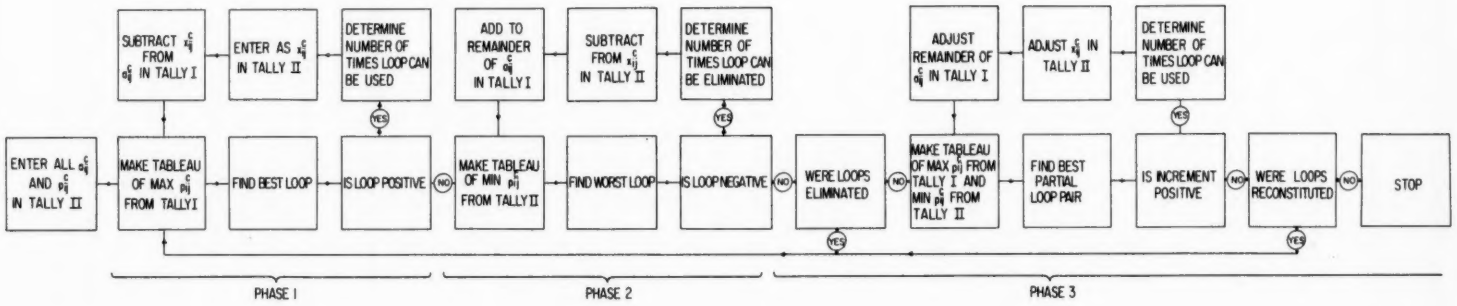

ENTER ALL a_{ij}^c AND p_{ij}^c IN TALLY II
MAKE TABLEAU OF MAX p_{ij}^c FROM TALLY I
SUBTRACT x_{ij}^c FROM a_{ij}^c IN TALLY I
ENTER AS x_{ij}^c IN TALLY II
DETERMINE NUMBER OF TIMES LOOP CAN BE USED
FIND BEST LOOP
IS LOOP POSITIVE
YES
NO
MAKE TABLEAU OF MIN p_{ij} FROM TALLY II
ADD TO REMAINDER OF a_{ij}^c IN TALLY I
FIND WORST LOOP
SUBTRACT FROM x_{ij}^c IN TALLY II
IS LOOP NEGATIVE
DETERMINE NUMBER OF TIMES LOOP CAN BE ELIMINATED
WERE LOOPS ELIMINATED
MAKE TABLEAU OF MAX p_{ij} FROM TALLY I AND MIN p_{ij}^c FROM TALLY II
ADJUST REMAINDER OF a_{ij}^c IN TALLY I
FIND BEST PARTIAL LOOP PAIR
ADJUST x_{ij}^c IN TALLY II
DETERMINE NUMBER OF TIMES LOOP CAN BE USED
IS INCREMENT POSITIVE
WERE LOOPS RECONSTITUTED
STOP
PHASE 1
PHASE 2
PHASE 3

Figure 1 - Abridged procedure for determining container schedules

x_{ij}^{c} = number of containers of commodity c shipped,

x_{ij}^{e} = number of empty containers shipped.

The a's and x's are nonnegative and

$$a_{ij}^{c} \gg x_{ij}^{c}.$$

Maximize

$$\sum_{i,j,c} x_{ij}^{c}\ p_{ij}^{c} - \sum_{i,j} x_{ij}^{e}\ p_{ij}^{e}.$$

Because detailed discussions of the linear programming model and the algorithm are included in [3] and [4], only the broad outline of the algorithm will be presented. The flow diagram shown in Figure 1 is a convenient way to trace out the procedure. However, it should not be interpreted as a program for digital computers. The algorithm has been developed as a pencil-and-paper process.

The flow diagram may be broken down into three phases. The inputs are the amounts of cargo to be shipped and the savings indices in terms of dollars per container. The basic operation is the construction of a tableau, such as shown in Figure 2, and the determination of "loops." A loop is a routing for a single container from port i back to port i, with one or more intermediate stops. For example, Figure 2 is laid out for a five-port trade route. The values in the table are the largest values of the savings index for the cargoes to be shipped from each port to every other port. By making a few trials, it can be determined that loop 1-2-4-1 produces the maximum savings or profits (in this case, \$145 per container). The total number of containers which can be shipped over this route depends upon the amount of cargo available for shipment. Note that the procedure is iterative. Each time the best loop is found for the cargo remaining to be shipped the amount of savings becomes less for each succeeding loop. This process is continued until no additional savings can be realized, thus completing Phase 1. Tally II in Figure 1 is a running list of all cargo to be containerized. Tally I starts as a list of all cargo to be shipped. As cargo is selected for containerization, it is removed from the Tally I; thus making the final tally a list of noncontainerized cargo.

Unfortunately, one cannot be certain that the optimum allocation has been found. In fact, it is most probable that the savings can be increased by eliminating those segments of the loops which have negative savings indices, if this can be done without destroying the balance of containers. The process of reconstituting loops and eliminating those which are negative because one or more of the segments are negative is given in Phase 2 of Figure 1.

Still further improvement can be realized by substituting one series of shipments in place of another, whenever such rescheduling produces more revenue. An example of such a "partial loop pair" is shown in Figure 3. The upper number in each cell is the largest saving index for the unassigned cargo still available for shipment, and the lower number is the smallest savings index for the cargo which had been designated for containers. In the case shown, a shipment of other cargo from port 2 to port 5, in place of the already "earmarked" shipments

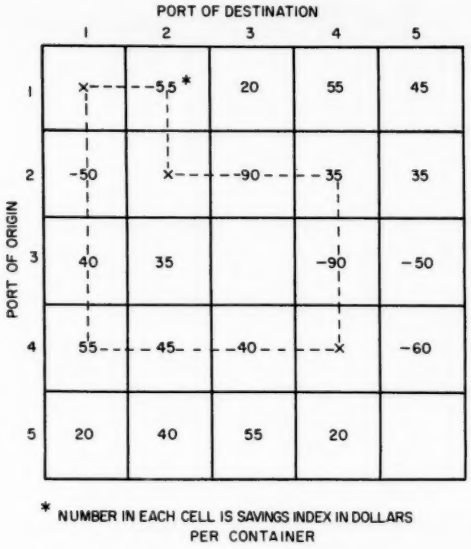


Figure 2 - Layout for determining loops in Phase 1

from ports 2 to 3 to 4 to 5, would result in a savings of $20. The determination of the best partial loop pairs is Phase 3, (Figure 1), the final phase of the computation.

It should be noted that the three phases are not equally significant. The major portion of the scheduling is accomplished by Phase 1. Phases 2 and 3 may produce only minor improvements, but they do lead to the achievement of an optimum schedule.

SUMMARY

A scheme for scheduling containers has been presented. No attempt has been made to include the details. They are available in the reports which have been cited. The procedure provides an insight into the movement of containers in a trade route, as well as producing a schedule. It has been noted that the computations are based upon cost data which are not readily available; however, these data could be obtained routinely by making minor changes in the present system. In view of the traditional nature of present-day maritime industry practices, the recognition that such data are necessary is perhaps more important than the details of the scheduling procedure.

REFERENCES

[1] Carrabino, J. D. "An Engineering Analysis of Cargo Handling, VI; Containerization," Report 57-56, July 1957, Dept. of Engr., Univ. of Calif., Los Angeles, California.

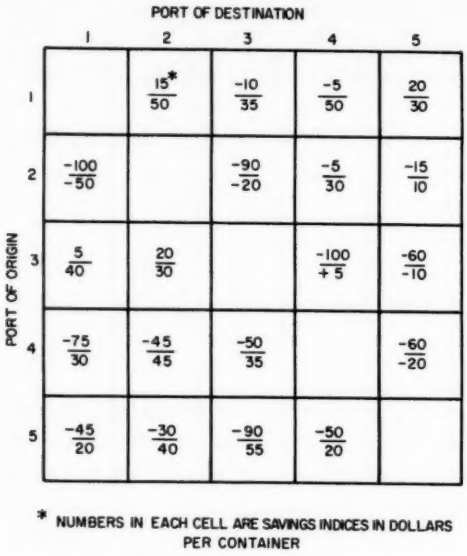


Figure 3 - Layout for determining partial loop pairs in Phase 3

[2] Carrabino, J. D. "Determination of Optimum Sizes and Economic Feasibility of Shipping Containers Using Operations Analysis Techniques," Journal of Industrial Engineering, Vol. 9, No. 6, pp. 513-520, November-December 1958.

[3] Toscano, F. C. "Allocation of Containers in a Trade Route," Research Techniques in Maritime Research, National Academy of Sciences, National Research Council Publication 720, 1959, pp. 52-88.

[4] Tannas, L. E. Technical Study in Cargo Handling, VIII; Procedure for Incorporating Containers Into a General Trade Route, Report 60-12, March 1960, Dept. of Engineering, University of California, Los Angeles, California.

* * *

RECENT DEVELOPMENTS IN THE JOB-SHOP SCHEDULING PROBLEM*

G. L. Thompson

Carnegie Institute of Technology

The job-shop scheduling problem is a large-scale combinatorial problem of great practical significance. First, a discussion of the size of the combinatorial problem involved will be given. Then, the various methods of attacks for solving this problem will be discussed and criticized by the use of examples. These methods include principally the integral linear programming, the heuristic, the loading rule, the simulation, and the Monte Carlo methods. The latter method, in the form devised by Giffler and Thompson, will be described in detail, and some computational results will be presented.

INTRODUCTION

The job-shop scheduling problem has existed since the innovation of large multi-machine factories. The fact that the operators of such factories have given the problem little attention is explained by the fact that reasonable-sized problems of this nature are essentially unsolvable by human efforts alone. The existence of large-scale computing machines now makes possible the consideration of this from a practical computational point of view.

A reasonably comprehensive review of efforts to solve the problem up to 1957 is given in [8]. In the present paper I shall briefly discuss and compare some of the current efforts on the problem. These include the loading rule, the simulation, the heuristic, the integral linear programming, and the Monte Carlo methods. I shall discuss very briefly the first three methods and then go into the last two in more detail.

DEFINITION OF THE PROBLEM

In the scheduling problem (the open problem of [5]), there are m commodities to be produced and n facilities which perform operations on them during their production. We assume that the order in which the commodities are to be processed by the machines is completely specified. This can be done most easily be means of a matrix called the facility sequence matrix, two examples of which are given in Figures 1 and 2.

C_1	3	1	2	4	6
C_2	2	3	5	6	1
C_3	3	4	6	1	2
C_4	2	1	3	4	5
C_5	3	2	5	6	1
C_6	2	4	6	1	5

Figure 1

C_1	1^5	2^2	3^8	4^7
C_2	3^8	1^4	4^5	2^3
C_3	4^6	1^7		

Figure 2

*This work was supported by the Office of Naval Research through a grant to the Graduate School of Industrial Administration of the Carnegie Institute of Technology.

The indices in the matrix of Figure 1 are those of the facilities that are to operate on the given commodity, and the order in which these indices appear is the order in which the commodity is to be "routed" to the various facilities. It is assumed here that all job times are 1. In Figure 2, the indices in the "matrix" are again the indices of the facilities in the order in which the commodity is to be processed. The superscripts of these indices are the corresponding job times.

Feasible schedules are most easily specified by means of a chart called a Gantt chart. For instance, a feasible schedule for the problem shown in Figure 1 is given in Figure 3. The entries in the chart are the indices of commodities, and these appear in the order in which each

F_1	-	-	4	1	3	6	2	5
F_2	6	4	2	5	1	3	-	-
F_3	5	3	1	2	4	-	-	-
F_4	-	6	3	-	-	1	4	-
F_5	-	-	-	-	2	5	6	4
F_6	-	-	6	3	-	2	5	1

Figure 3

machine processes them. Idle times are represented by dashes. Similarly, a feasible schedule for the problem given in Figure 2 is given in Figure 4. Here the repetition of an index or dash indicates the number of time periods that the commodity with that index or idle time are on the machine. Later we shall see that each of these schedules is an optimal feasible schedule for its corresponding problem.

F_1	1	1	1	1	1	-	-	-	2	2	2	2	3	3	3	3	3	3	-	-	-	-	-	
F_2	-	-	-	-	-	1	1	-	-	-	-	-	-	-	-	-	-	-	2	2	2	-	-	-	-	-
F_3	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	-	-	-	-	-	-	-	-	-	-
F_4	3	3	3	3	3	3	-	-	-	-	-	-	2	2	2	2	2	1	1	1	1	1	1	1

Figure 4

THE LOADING RULE, SIMULATION, AND HEURISTIC METHODS

Loading rules, as usually discussed, can be characterized as local rules applied by the operators of machines to decide which of a large number of commodities waiting to be processed on the machine should be worked on next. For instance, an operator may choose to work on the job that arrived first, or the job he can complete most quickly, or the job that has the maximum number of uncompleted operations still to be performed, etc. He may even just choose a job at random from among those waiting.

Usually these rules are described as if they lead to a unique set of decisions. But this does not follow whenever ambiguity of choice arises where two or more jobs waiting to be done

satisfy the criterion simultaneously. For instance, if the operator can complete five different jobs all in the same length of time, which one should he work on next? It might seem as if it wouldn't make any difference, and he could just choose at random. But, in fact, if he chooses at random, then the net effect of all such random choices by all the operators is equivalent to choosing at random from among the set of all possible schedules that can be generated in this way. These do not all necessarily have the same length (see the next section) and, hence, if the random choice among all such global schedules were made a random number of times, and the one with the shortest length were selected for use, an average saving of total completion time of several percent could be effected.

Simulation techniques are directed toward the study of an actual manufacturing situation and the simulation of the behavior of various machines or machine groups in a plant. In this way one can investigate the effect of machine breakdowns, priority systems, changes in operating rules, etc. Reports of such efforts are given in [2], [3], and [4].

Heuristic rules are developed for production scheduling problems in order to develop "skillful" schedulers. In this case, skill, or expertise, has the same connotation as the skill of a chess master, or other proficient master of a game. Loading rules can be regarded as heuristic rules of not too great akill. The skillfulness of a heuristic rule can be measured against other heuristic rules, e.g. the loading rules. Work is being done on this approach by W. Gere and others at Carnegie Tech.

INTEGRAL LINEAR PROGRAMMING AND MONTE CARLO METHODS

There have been three formulations of the job-shop scheduling problem as an integral linear program: that of Wagner [9], of Bowman [1], and of Manne [7]. Bowman discusses his method in terms of the example in Figure 2, and I shall use that example to compare the three formulations. I shall not go into the explicit descriptions of the formulations but merely estimate the size of the resulting computational problem.

Bowman estimates that to formulate the problem of Figure 2 in his terms would require an integer linear programming problem containing 300 to 600 variables and many more constraints. The formulation by Wagner would involve possible fewer variables and constraints, but would be of substantially the same order of magnitude. Manne's formulation, the most compact of the three, would, by my estimates, require 31 variables and 94 constraints. This last formulation could probably be put on existing computing machines and solved in a reasonable length of time. It is to be emphasized, however, that in the solution of an integral linear programming problem, additional constraints are added during the course of the solution; so that the ultimate number of constraints will be larger.

None of these authors claims that his formulation is computationally practical, but they are theoretical formulations suitable for further study.

An entirely different approach was undertaken by Giffler and Thompson [5]. Their first effort was in devising an algorithm for the complete enumeration of all active feasible schedules. (They define an active feasible schedule to be one with the following properties: (a) no machine is idle for a length of time sufficient to completely process a simultaneously idle commodity, and (b) whenever an assignment of a commodity to a machine has been made, its processing is started at the earliest time that both the machine and the commodity are free.) In Figure 5 we produce all the computations needed to completely enumerate the five active

feasible schedules that exist for the problem in Figure 2. For a description of the algorithm used, see [5]. The schedules marked with an X are incomplete, since they contain conflicts.

F_1			F_2		F_3		F_4				
1	2	3	1	2	1	2	1	2	3		
5	12	13	7		15	8			6	1	X
5	12	13	7		16	8			6	11	X
5	27	13	7	35	15	23	22	32	6	12	L35
5	12	19	7		16	8	23	17	6	111	X
5	17	13	7		16	8	23	22	6	112	X
5	12	19	7	31	16	8	23	28	6	1111	L31
5	12	19	7	20	16	8	24	17	6	1112	L24
5	17	13	7	31	16	8	23	28	6	1121	L31
m5	17	13	7	25	16	8	29	22	6	1122	L29

Figure 5

Schedules derived from these by resolving conflicts are indicated in the Dewey decimal numbering system. Completed schedules are marked with an L Followed by a number indicating the completion time. The Gantt Chart in Figure 4 is the one corresponding to schedule 1112 in Figure 5, and is the unique active optimum schedule for the problem. There are 1889 inactive optimum schedules equivalent—in the sense of [5]—to this one.

One thing that is immediately clear from the above computation is that the narrowing of the search for optimum schedules to the search for active optimum schedules reduces the complete enumeration method of solving the scheduling problem to a computationally feasible method for small problems. Also, the computational effort in solving even the Manne integer linear programming formulation of the problem is many times greater than this complete enumeration method. The work displayed in Figure 5 can be completed with pencil and paper in less than 30 minutes and can be done on a large computer in a fraction of a second.

The complete enumeration of the 84.802 active feasible schedules of the problem of Figure 1 was carried out by an IBM 704 in 70 minutes and yielded the following distribution:

Completion Time	Number Found	Percentage
8	933	1.1
9	9333	11.0
10	24516	28.9
11	25987	30.6
12	13809	16.3
13	10224	12.1

The Gantt chart in Figure 3 shows one of the optimal schedules that completes in 8 time periods. Note that if one chooses just one schedule at random, the most probable length schedule that will be obtained is 11; which is 38 percent longer than the optimum.

This problem was the largest for which the complete enumeration method has been used. Much larger schedules (for instance one in which m = 200 and n = 10) have been computed by a Monte Carlo version of the complete enumeration method. These computations are reported more fully in [6].

The Monte Carlo method resolves conflicts at random when they occur. It does not guarantee the finding of an optimal schedule, but it does permit the computation of a fair number of feasible schedules chosen at random, and the shortest one found in the process can then be used. By continuing the process long enough, it is possible to make the probability of not having the optimum as small as desired.

The Monte Carlo method, if run for only a few choices of feasible schedules, qualifies merely as a skillful scheduler in the sense of the preceding section. It compares favorably with many of the commonly used loading rules—doing better than they do for some classes of problems, and worse in other cases. For further discussion, see [6].

CONCLUSION

It is clear that no computationally-feasible formulation of an integer linear programming model of the job-shop scheduling problem has as yet been developed. But, with improvements in the computations of integer programs and improved formulations of the problem, there is ample hope that progress will be made in that direction in the long run.

Progress toward approximate or skillful computational methods to solve scheduling problems, by the use of heuristic, Monte Carlo, or combinations of these methods, seems currently very promising. It is quite likely that the first practical applications of scheduling techniques will be made by use of these methods, or possibly by use of simulation techniques.

REFERENCES

[1] Bowman, E. H., "The Schedule Sequencing Problem," Operations Research, Vol. 7, 1959, pp. 621-624.

[2] Conway, R. W., Johnson, B. M., and Maxwell, W. L., "The Cornell Research Simulation," Report of the Cornell Industrial Engineering Department.

[3] _____, "An Experimental Investigation of Priority Dispatching," ibid.

[4] _____, "Some Problems of Digital Systems Simulation," Management Science, Vol. 6, 1959, pp. 92-110.

[5] Giffler, B., and Thompson, G. L., "Algorithms for Solving Production Scheduling Problems," to be published in Operations Research.

[6] Giffler, B., Thompson, G. L., and VanNess, V., "Numerical Experience With Algorithms for Solving Production Scheduling Problems," in preparation.

[7] Manne, A. S., "On the Job Shop Scheduling Problem," Cowles Foundation Discussion Paper, No. 73, May 1959.

[8] Sisson, Roger, L., "Methods of Sequencing in Job Shops—A review," Operations Research, Vol. 7, 1959, pp. 10-29.

[9] Wagner, H. M., "An integer linear-programming model for machine shop scheduling," Naval Research Logistics Quarterly, Vol. 6, 1959, pp. 131-140.

PANEL 6
INVENTORY CONTROL

SOME REMARKS ON BAYES SOLUTIONS TO THE INVENTORY PROBLEM*

Herbert E. Scarf

Stanford University

A functional equation is described for the inventory problem in which the demand distribution contains an unknown parameter with a known a priori distribution. If the demand distribution is a member of the exponential family, then the minimum cost is a function of two variables: current stock and past mean demand. Some properties of the optimum policies are described, and some conditions are given which imply that the functional equation can be reduced to one involving functions of one variable only. Under these conditions, the computation of optimal policies becomes no more difficult than the corresponding computation with a known demand distribution.

INTRODUCTION

In [2] we discussed a technique for the computation of optimal inventory levels in the case where the demand distribution contains an unknown statistical parameter. It was assumed that initially the parameter could be described by a Bayes distribution, which would be subsequently revised on the basis of additional demand information. By assuming that the demand distribution is a member of the exponential family, so that the cumulative observed demand is a sufficient statistic for the unknown parameter, it was shown that the optimal inventory levels could be obtained by the recursive computation of a sequence of functions of two variables. The first of these variables—as in the case of complete knowledge of the demand distribution—is the current stock level x; the second variable is s, the sum of the observed demands up to the present time.

Functions of two variables are, of course, difficult to compute recursively. We shall show, in this paper, that if several additional assumptions are made, then it is possible to determine the optimal levels by the recursive computation of functions of one variable alone (See Equations 5 and 6).

The first of these assumptions is that the purchase cost be proportional to the amount purchased, in other words, that there be no set up cost. This is undoubtedly a severe restriction. On the other hand, whatever applications are made of the technique described in this paper will most probably be to expensive, low-demand items; and for these items the set-up cost can frequently be incorporated in the unit cost.

*This work was performed while the author was a consultant for the Planning Research Corporation, on a contract for the Bureau of Supplies and Accounts, Department of the Navy.

The second assumption will be that the demand distribution be a member of the Γ family, that is

$$\phi(x|w) = \frac{w(wx)^{a-1} e^{-wx}}{\Gamma(a)}, \tag{1}$$

with w as the unknown parameter and a as a fixed constant. (For this family of distributions, the ratio of mean to standard deviation is $\sqrt{a}$. Therefore, all possible means are permitted, but the standard deviation is to be in a fixed ratio to the mean.) We shall assume moreover that the a priori distribution of w also is from the Γ family.

$$f(w) = \frac{\lambda(\lambda w)^{b-1} e^{-\lambda w}}{\Gamma(b)}, \tag{2}$$

with the constants λ and b known.

The one remaining assumption is that the holding cost is proportional to the excess of stock and the shortage cost proportional to the deficit of stock.

The main idea of the argument presented in this paper is that if we make the above assumptions the optimal policies will be given by single critical numbers which are actually proportional to λ plus cumulative past demand. The only quantities to be computed are the proportionality constants, and these may be obtained by techniques which are virtually identical to those used in the case of a known demand distribution.

THE COMPUTATION TECHNIQUE

Let us begin by an examination of the functional equation derived in [2]. In this paper we shall make the same stipulation as in [2] concerning the length of the time-lag in delivery; that is, we shall take it to be zero. If there is a time-lag in delivery, and excess demand is backlogged, then the procedure described in Chapter 9 of [1] may be used in conjunction with the method described here to obtain a recursive computation based on functions of a single variable. We define a sequence of functions $C_n(x, s)$ to be the minimum, discounted, expected cost for the remainder of the program, if at the beginning of the n^{th} period the current stock level is x, and the cumulative past observed demand is s. The functional equation is, as usual, obtained by examining the cost consequences of a purchasing decision for the present period and for the remainder of the program.

If we decide to raise our stock from x to y units, then we incur a cost of c(y-x), where c is the unit cost. We must then take account of the expected holding and shortage costs at the end of the present period. If the demand during the period is ξ, then these costs are given by the expectation of

$$h \max(0, y - \xi) + p \max(0, \xi - y)$$

This expectation is to be taken with respect to the a posteriori distribution of demand in the present period when given that the total past demand was s. It is a simple calculation to show that the a posteriori distribution is given by:

$$\phi_n(x|s) = \frac{\Gamma(na + b)}{\Gamma(a)\Gamma[(n-1)\,a+b]} \frac{\frac{1}{(s+\lambda)}\left(\frac{x}{s+\lambda}\right)^{a-1}}{\left(1 + \frac{x}{s+\lambda}\right)^{na+b}},$$

and therefore the expected holding and shortage costs are given by

$$L_n(y|s) = h\int_0^y \left(y - \xi\right)\phi_n\left(\xi\,|s\right)d\,\xi + p\int_y^\infty \left(\xi - y\right)\phi_n\left(\xi|s\right)d\,\xi,$$

for $y \geq 0$ and

$$p\int_0^\infty \left(\xi - y\right)\phi_n\left(\xi|s\right)d\,\xi \qquad \text{for } y \leq 0\,.$$

The state of the system at the beginning of the next period, if the demand during this period is ξ, is given by $(y - \xi,\ s + \xi)$, and therefore the discounted future costs will be

$$\alpha\int_0^\infty C_{n+1}\left(y - \xi\,,\ s + \xi\right)\phi_n\left(\xi|s\right)d\,\xi\,.$$

These considerations lead to the following functional equation:

$$(3)\quad C_n(x,s) = \min_{y \geq x}\left[c(y - x) + L_n(y \mid s) + \alpha\int_0^\infty C_{n+1}\left(y - \xi\,,\ s + \xi\right)\phi_n\left(\xi|s\right)d\,\xi\right],$$

with $C_N \equiv 0$ if the program lasts a total of N periods.

The important fact to be used in simplifying this equation is that

$$\phi_n(x \mid s) = \frac{1}{(s+\lambda)}\phi_n\left(\frac{x}{s+\lambda}\right),$$

where $\phi_n(x)$ is defined to be

$$\phi_n(x) = \frac{\Gamma(na + b)}{\Gamma(a)\Gamma[(n-1)\,a+b)]}\ \frac{x^{a-1}}{(1+x)^{na+b}}$$

Let us examine the consequences of this relation for the function $L_n(y \mid s)$. We have

$$h \int_0^y \left(y - \xi\right) \phi_n(\xi|s)\, d\xi = \frac{h}{s+\lambda} \int_0^y (y - \xi)\, \phi_n\left(\frac{\xi}{s+\lambda}\right) d\xi \quad .$$

Make the change of variable $\xi = (s + \lambda)t$ and we obtain

$$h \int_0^{\frac{y}{s+\lambda}} [y - (s+\lambda)t]\, \phi_n(t)\, dt = (s+\lambda) h \int_0^{\frac{y}{s+\lambda}} \left(\frac{y}{s+\lambda} - t\right) \phi_n(t)\, dt.$$

We do the same sort of thing to the shortage cost component, and we see that

$$L_n(y \mid s) = (s + \lambda)\, L_n\left(\frac{y}{s+\lambda}\right),$$

where

$$L_n(y) = h \int_0^y (y - \xi)\phi_n(\xi)\, d\xi + p \int_y^\infty (\xi - y)\phi_n(\xi)\, d\xi \ ,$$

for $y \geq 0$.

This relationship tells us that the expected holding and shortage cost function can very well be represented by a function of a single variable. It seems intuitively clear that a similar sort of representation is valid for the cost functions $C_n(x, s)$ and, as we shall show, this is indeed the case. Specifically, we shall show that

$$C_n(x, s) = (s + \lambda)\, C_n\left(\frac{x}{s+\lambda}\right) \ , \tag{4}$$

where C_n of a single variable is a new function satisfying a functional equation that we shall indicate. The proof of Equation 4 is by induction backwards on n. Let us assume that this is true for $(n + 1)$. Then

$$C_n(x, s) = \min_{y \geq x} \left[c(y - x) + (s + \lambda)\, L_n\left(\frac{y}{s+\lambda}\right) + \alpha \int_0^\infty C_{n+1}(y - \xi\ , s + \xi)\, \phi_n(\xi|s)\, d\xi \right].$$

The last term is equal (by the induction assumption) to

$$\alpha \int_0^\infty C_{n+1}\left(\frac{y - \xi}{s + \xi + \lambda}\right)(s + \xi + \lambda)\,\phi_n(\xi|s)\,d\xi$$

$$= \alpha \int_0^\infty C_{n+1}\left(\frac{y - \xi}{s + \xi + \lambda}\right)\frac{(s + \xi + \lambda)}{s + \lambda}\,\phi_n\left(\frac{\xi}{s + \lambda}\right)d\xi .$$

Now make a change of variable $\xi = (s + \lambda)t$, and we obtain

$$(s + \lambda)\,\alpha \int_0^\infty C_{n+1}\left(\frac{\frac{y}{s + \lambda} - t}{1 + t}\right)(1 + t)\,\phi_n(t)\,dt .$$

Inserting this in the functional equation, we obtain

$$C_n(x, s) = (s + \lambda) \min_{\frac{y}{s+\lambda} > \frac{x}{s+\lambda}} \left[c\left(\frac{y}{s + \lambda} - \frac{x}{s + \lambda}\right) + L_n\left(\frac{x}{s + \lambda}\right) \right.$$

$$\left. + \alpha \int_0^\infty C_{n+1}\left(\frac{\frac{y}{s + \lambda} - t}{1 + t}\right)(1 + t)\,\phi_n(t)\,dt \right] .$$

Since we also know in [2] that the optimal policy is defined by a single critical level for each s, we conclude that there exists a number a_n, with the property that the optimal policy is

(5)
$$\begin{cases} \text{if } x \le a_n(\lambda + s), \text{ select } y = a_n(\lambda + s) \\ \text{if } x \ge a_n(\lambda + s), \text{ select } y = x \end{cases}$$

It is also clear that if we define $C_n(x)$ by the equation

(6)
$$C_n(x) = \min_{y \ge x} \left[c(y - x) + L_n(y) + \alpha \int_0^\infty C_{n+1}\left(\frac{y - t}{1 + t}\right)(1 + t)\,\phi_n(t)\,dt \right].$$

then $C_n(x, s) = (s + \lambda)\,C_n\left(\frac{x}{s + \lambda}\right)$ and the induction step is complete.

REFERENCES

[1] Karlin, S., and Scarf, H., "Inventory Models of the Arrow-Harris-Marschak Type with Time Lags," Chapter 9, in Studies in the Mathematical Theory of Inventory and Production, Kenneth J. Arrow, S. Karlin, and H. Scarf, Stanford University Press, Stanford, 1958.

[2] Scarf, H., "Bayes Solutions of the Statistical Inventory Problem," Annals of Mathematical Statistics, Vol. 30, 1959, pp. 490-508.

* * *

REPLACEMENT OF PERIODICALLY INSPECTED EQUIPMENT*

(An Optimal Optional Stopping Rule)

C. Derman and J. Sacks

Columbia University

The problem of choosing an optimal replacement rule is considered for deteriorating equipment where the amount of deterioration is observed periodically. A "near" optimal rule is obtained. The main tool is an optimal stopping lemma related to semimartingales.

INTRODUCTION

For orientation, consider the following elementary replacement problem. Suppose a piece of equipment in operation has a random service life Z with distribution function G. If the piece is replaced before it fails, a cost c is incurred; if it fails before it is replaced, then the piece is immediately replaced at a cost of c + A (i.e., a penalty cost A is incurred).

For any replacement rule R, and i = 1, 2, . . . , let C_{Ri} be the cost (i.e., c or c+ A) associated with the i^{th} replacement cycle and let N_{Ri} be the length of the i^{th} replacement cycle. One criterion for evaluating the effectiveness of the rule R is the average cost per unit time,[1]

$$\phi_R = \lim_{T\to\infty} \frac{\sum_{i=1}^{M_R(T)} C_{Ri}}{T} = \frac{E\,C_R}{E\,N_R} \tag{1}$$

where $M_R(T)$ is the number of replacements in the interval (0, T).

However, for some situations, the criterion

$$\phi = E\left(\frac{C}{N}\right)$$

is appropriate.[2]

Now, if we restrict the class of possible rules to those of the form

$$N = \min(t, Z)$$

(i.e., replace at time t if failure has not already occurred)

*Research sponsored by the Office of Naval Research.

[1]We have tacitly assumed that the replacement rule is the same for each cycle so the $\{C_{Ri}\}$ and $\{N_{Ri}\}$ are sequences of identically and independently distributed random variables. Consequently, the strong law of large numbers guarantees (1).

[2]Here, and henceforth, the subscript R will be omitted whenever no confusion will result

where t is a constant, then

$$\phi = \frac{c + A\,G(t)}{\int_0^t s\,d\,G(s) + t[1 - G(t)]}$$

and

$$\phi' = (c + A) \int_0^t \frac{1}{s}\,d\,G(s) + \frac{c}{t}\,[1 - G(t)].$$

If G is known, then ϕ and ϕ' can be minimized, at least in principle, by choosing t appropriately. Such minimizing values of t then provide optimal rules among the class considered.

This and similar replacement problems are discussed by Klein and Rosenberg [4].

We are concerned with the following modified problem which involves the efficient use of acquired information. Suppose the equipment is such that it has a measureable characteristic associated with it which is decreasing in time; that is, the equipment deteriorates. We shall assume that at times $t = 1, 2, \ldots$, the amount of deterioration X_t occurring during the interval $(t - 1, t)$ is observable and that the sequence $\{X_t\}$ $t = 1, 2, \ldots$ is a sequence of independent, identically distributed, nonnegative, random variables with common distribution function F. The equipment is known to fail if $\sum_{t=1}^{n} X_t > L$ (a given constant). Hence Z, the service life, satisfies the inequalities $k - 1 < Z \leq k$, if k is the smallest integer n, such that $\sum_{t=1}^{n} X_t > L$. For convenience, we shall also assume that the equipment can be replaced only at the time points $t = 1, 2, \ldots$. The cost structure is, as before with c, the cost of a replacement before failure, and $c + A$ the cost of a replacement after failure.

Rather than restrict ourselves to replacement rules of the type mentioned earlier, suppose we consider the class ζ of rules of the form

$$N = \min(\nu, [Z] + 1),$$

where $[Z]$ denotes the greatest integer less than Z and ν is an optional stopping rule (i.e., conditions for determining $\nu = s$ $(s = 1, \ldots)$ are in terms of $X_1, \ldots, X_s$). This means that we allow the rules to depend on the observed deterioration. We restrict ν, and consequently N, to integer values. The problems are to find rules R^* and R_0 in ζ such that

$$\phi_{R^*} = \min_{R \in \zeta} \phi_R$$

and

$$\phi_{R_0} = \min_{R \in \zeta} \phi'_R$$

There is a suggested method for deriving R^* given in Arrow, Karlin, and Scarf[3] ([1], p. 283). Namely, consider a discounted cost criterion (as in inventory problems) and use the functional equation approach to derive an optimal rule R_α, where α is the discount factor; and let $\alpha \longrightarrow 1$. Presumably, the necessary interchange of limits is justifiable in this case.

We shall not attempt to exploit this method here. Instead we shall provide machinery for finding R_0. This machinery is contained in the lemma of the next section.

AN OPTIMAL OPTIONAL STOPPING LEMMA

Let $\{\mathfrak{F}_n,\ n = 1,2,\ldots\}$ be a sequence of σ - fields of a sample space Ω with $\mathfrak{F}_n \subset \mathfrak{F}_{n+1}$ $n = 1,2,\ldots$. Let $\{Y_n\}$ be a sequence of random variables with Y_n measurable with respect to $\mathfrak{F}_n$ and such that $E\,Y_n$ exists and is finite for all n. Let C be the class of all optional stopping rules such that $E\,N < \infty$.

In order to understand more quickly the lemma below, it will help to think of a sequence of gambling games (not necessarily independent) where Y_n denotes the winnings of a player after the n^{th} game. The statement, "Y_n measurable with respect to $\mathfrak{F}_n$" is then interpreted as meaning that the winnings after the n^{th} game are dependent only on the first n games (i.e., the σ-fields $\mathfrak{F}_n$ are generated by conditions on the first n games). An optimal stopping rule is a rule for telling the player when to stop. An optimal rule is one that maximizes the expected winnings.

LEMMA:[4] If there exists an optional stopping rule N^* with:

(a) $E\,N^* < \infty$

(b) $E\{Y_n \mid \mathfrak{F}_{n-1}\} \geq Y_{n-1}$ when $n \leq N^*(\omega)$

$\leq Y_{n-1}$ $n > N^*(\omega)$

for almost all $\omega \in \Omega$; and if there is some $M < \infty$ such that

(c) for all n, $E\{|Y_{n+1} - Y_n| \mid \mathfrak{F}_n\} \leq M$,

then

$$E\,Y_{N^*} = \operatorname*{Max}_{N \in C} E\,Y_N. \tag{2}$$

[3] We are indebted to Professor Herbert Scarf for an informative conversation regarding this approach.

[4] Y. S. Chow has since obtained stronger results along these lines.

PROOF: The proof is, with simple modifications, that given by Doob ([2], Theorem 2.2, Chapter VII); such an argument being appropriate since condition (b) gives a kind of semi-martingale structure to $\{Y_n\}$. We show first that if $N \in C$, then $E|Y_N| < \infty$.

Let $\xi_1 = |Y_1|$, $\xi_t = |Y_t - Y_{t-1}|$ if $t > 1$.

Then $|Y_t| \leq \xi_1 + \ldots + \xi_t$ and, using (c),

$$E|Y_N| = \sum_{k=1}^{\infty} \int_{N=k} |Y_k| \, dP \leq \sum_{k=1}^{\infty} \sum_{i=1}^{k} \int_{N=k} \xi_i \, dP = \sum_{i=1}^{\infty} \sum_{k=i}^{\infty} \int_{N=k} \xi_i \, dP$$

$$= \sum_{i=1}^{\infty} \int_{N \geq i} \xi_i \, dP \leq M \sum_{i=1}^{\infty} P\{N \geq i\} = ME(N) < \infty .$$

Thus (2) makes sense. We now show

(3) $$\int_{N<N^*} Y_N \, dP \leq \int_{N<N^*} Y_{N^*} \, dP .$$

By the first part of (b),

(4) $$\int_{N<N^*} Y_N \, dP = \sum_{k=1}^{\infty} \int_{\substack{N^*>k \\ N=k}} Y_k \, dP \leq \sum_{k=1}^{\infty} \int_{\substack{N^*>k \\ N=k}} Y_{k+1} \, dP$$

$$= \int_{N+1=N^*} Y_{N^*} \, dP + \int_{N+1<N^*} Y_{N+1} \, dP .$$

Continuing in the same fashion we obtain, for any r,

(5) $$\int_{N<N^*} Y_N \, dP \leq \sum_{j=1}^{r} \int_{N+j=N^*} Y_{N^*} \, dP + \int_{N+r<N^*} Y_{N+r} \, dP$$

as $r \to \infty$ the first term on the right side of (5) converges to

$$\int_{N<N^*} Y_{N^*} \, dP ,$$

and if we show

$$\lim_{r\to\infty} \int_{N+r\leqq N^*} |Y_{N+r}|\, dP = 0 \tag{6}$$

we are finished. To obtain (6), let $z_{r+1} = |Y_{r+1}|$, $z_t = |Y_t - Y_{t-1}|$ $t > r+1$. Then, using (c) towards the end,

$$\int_{N+r\leqq N^*} |Y_{N+r}|\, dP = \sum_{k=1}^{\infty}\sum_{j=r+k}^{\infty} \int_{\substack{N=k\\N^*=j}} |Y_{k+r}|\, dP \tag{7}$$

$$\leqq \sum_{k=1}^{\infty}\sum_{j=r+k}^{\infty}\sum_{i=r+1}^{k+r} \int_{\substack{N=k\\N^*=j}} z_i\, dP$$

$$= \sum_{i=r+1}^{\infty}\sum_{j=1}^{\infty}\sum_{k=i-r}^{j-r} \int_{\substack{N=k\\N^*=j}} z_i\, dP$$

$$\leqq \sum_{i=r+1}^{\infty}\sum_{j=i}^{\infty} \int_{N^*=j} z_i\, dP$$

$$= \sum_{i=r+1}^{\infty} \int_{N^*\geqq i} z_i\, dP$$

$$\leqq M \sum_{i=r+2}^{\infty} P\{N^*\geqq i\} + \int_{N^*\geqq r+1} |Y_{r+1}|\, dP .$$

As $r\to\infty$ the first term of the last inequality in (7) goes to 0 because $EN^* < \infty$. Since

$$\int_{N^*\geqq r+1} |Y_{r+1}|\, dP \leqq \int_{N^*\geqq r+1} \left(\xi_1 + \ldots + \xi_{N^*}\right) dP \tag{8}$$

and
$E\left(\xi_1 + \ldots + \xi_{N^*}\right) < \infty$ (as we have shown in the first part of the proof), it follows that, as $r\to\infty$, the right side of (8) goes to 0. Thus (6) is established and so is (3).

We now show that

$$\int_{N>N^*} Y_N \, dP \leqq \int_{N>N^*} Y_{N^*} \, dP . \tag{9}$$

The proof of (9) is the same as the proof of (3) by interchanging the roles of N and N^* in the proof of (3), and observing that it is the second half of (b) which is relevant. Formulas (3) and (9) put together, show that

$$E\, Y_{N^*} \geqq E\, Y_N$$

which yields (2).

REMARKS

I. Other conditions than (a) and (c) can be used to give the same result as (2), namely conditions such as cited in [2].

II. Interchanging the inequalities in (b) (i.e., assuming $E\{Y_n \mid \mathfrak{F}_{n-1}\} \leqq Y_{n-1}$ when $n \leqq N^*(\omega)$, etc.) provides us with a rule N^* which minimizes $E\, Y_N$.

III. It is a temptation to pin on the lemma the label "intuitively obvious" because of condition (b). However, by relaxing conditions (a) and (c), it is not difficult to construct counterexamples for which the conclusion of the lemma does not hold. It is exactly this point which lends interest to the problem of finding R_0.

IV. We shall apply the lemma to our problem in section 3. However, one can see that the lemma will apply in a variety of situations. For example, the following problem was considered by G. Schwarz and also by H. Robbins, who obtained the solution below by other means.

Let $c > 0$ and let $\{X_n\}$ be a sequence of independent and identically distributed random variables with common distribution function F having a first moment. Let[5]

$$Y_n = \max(S_1, \ldots, X_n) - cn.$$

The problem is to find an optional stopping rule N which maximizes $E\, Y_N$. Let N^* be the first integer n such that $\max(X_1, \ldots, X_n) \geqq a$ where a is such that

$$\int_a^\infty (x - a)\, dF(x) - c = 0.$$

It is easy to check that N^* satisfies (a) and (b) and that $\{Y_n\}$ satisfies (c) where $\mathfrak{F}_n$ is the σ-field generated by $(X_1, \ldots, X_n)$.

[5] Schwarz considered $Y_n = X_n - cn$. However, the two sequences are essentially equivalent with respect to the problem under consideration.

APPLICATION OF THE LEMMA

In order to apply the lemma to the replacement problem, let

$$Y_n = C/N \qquad \text{if } n < Z$$

$$= \frac{c + A}{[Z] + 1} \qquad n \geq Z.$$

If N is the number of time periods until a replacement when rule R is used, then $Y_N = C/N$. We are, therefore, interested in choosing N to minimize $E\,Y_N$ (see II under remarks,).

Since the Y_n are uniformly bounded, condition (c) of the lemma holds. Letting $S_n = \sum_{t=1}^{n} X_t$ $(n = 1, 2, \ldots)$, we see that

$$E\,(Y_n \mid X_1, \ldots, X_{n-1}) = \frac{c}{n} + \frac{A}{n}\left[(1 - F\,(L - S_{n-1})\right] \text{ if } n - 1 < Z$$

$$= \frac{c + A}{[Z] + 1} \text{ if } n - 1 \geq Z.$$

Let $\mathfrak{Z}_n$ be the σ-field generated by conditions on $X_1, \ldots, X_n$.
Then

$$E\left(Y_n \,\middle|\, \mathfrak{Z}_{n-1}\right) \leq \frac{c}{n-1} = Y_{n-1} \text{ if } n-1 < Z \text{ and } 1 - F\,(L - S_{n-1}) < \frac{c}{A\,(n-1)}$$

$$\geq \frac{c}{n-1} = Y_{n-1} \text{ if } n-1 < Z \text{ and } 1 - F\,(L - S_{n-1}) \geq \frac{c}{A(n-1)}$$

$$= \frac{c + A}{[Z] + 1} = Y_{n-1} \text{ if } n-1 \geq Z.$$

Let n^* be the smallest n such that

$$1 - F\,(L - S_{n-1}) \geq \frac{c}{A\,(n-1)}.$$

Then, for all $n > n^*$, $1 - F(L - S_{n-1}) \geq c/A\,(n-1)$, since $1 - F(L - S_{n-1})$ is increasing in n and $c/A\,(n-1)$ is decreasing in n. Let $N^* = \min(n^* - 1, [Z] + 1)$; then routine computation shows that condition (b) of the lemma is satisfied with the inequalities reversed. Ruling out the degenerate case where $P\,(X = 0) = 1$, we have, from renewal theory, that $E\,N^* \leq 1 + E\,Z > \infty$; thus, condition (a) is satisfied. Similarly, all replacement rules (stopping rules) in ζ have finite expectations. Thus, all the conditions of the lemma (see remark II) are satisfied, and it follows that R_0 is given by the rule: replace when the equipment has been in service N^* time periods.

THE CASE WHERE F(x) IS UNKNOWN

The rule R_0, defined in section 3, depends on the distribution function F(x). It is, namely, replace at the first n such that

$$S_n \geq a_n = \max\left[L - F^{-1}\left(1 - \frac{c}{An}\right),\ 0\right]$$

for $n = 1, 2, \ldots$ where a_n is taken as the smallest such value when not uniquely defined by the above expression.

If F(x) is unknown, it seems reasonable to substitute for F(x) the empirical distrubution function (or any other estimate of F(x) which converges to F(x) with probability one) based upon the deterioration observed in all the preceding cycles. If such a substitute rule R'_0 is used, it is then a question as to what we can say about its performance characteristic.

Before considering this question, we digress here to consider a more general question. Very often, it is possible to derive, as in our replacement problem, an optimal rule R of behavior, if an appropriate distribution function H exists and is known. Let W_i be the value of the random variable which denotes the cost associated with the i^{th} application of the rule. Then, because the rule is applied repeatedly, we can consider

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} W_i = EW$$

which exists with probability one, due to the fact that all W'_i are independent and identically distributed random variables (we have assumed $E|W| < \infty$). Since R is optimal, we then know that the time average of the associated costs (being the same as the expected value) is also minimized.

Suppose, however, that initially H is not known; that on each application of any rule, observations are obtained on H; that the observations can be used to compute an estimate H_i of H; that, as $i \to \infty$, H_i converges to H with probability one. Let R' be the rule defined as follows: Use, on the $(i+1)^{th}$ application, the optimizing procedure derived by considering H_i as the true distribution. Let W'_i denote the cost associated with the i^{th} application of R'. Now, in a number of cases we will have $\lim_{i\to\infty} E(W'_{i+1}|H_i) = EW$ with probability one (see Johns [3] and Robbins [6] for related ideas). This convergence, when it exists, justifies, in a sense, the rule R'. However, it does not follow that

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} W'_i}{n} = EW \tag{10}$$

with probability one, since all W'_i are not necessarily independent and identically distributed. Thus, if the time average criterion is of interest, it is also important to know when (10) holds.

The theorem below is an initial result in this direction. Let $\mathcal{F}_i$ $(i = 1, 2, \ldots)$ be a sequence of σ-fields on a sample space Ω with $\mathcal{F}_i \subset \mathcal{F}_{i+1}$ for all i. Suppose the sequence $\{W'_i\}$

is such that W'_i is measurable with respect to $\mathfrak{Z}_i$ for each i,

$$E\left\{W'_{i+1} \mid \mathfrak{Z}_i\right\} = E\left\{W'_{i+1} \mid H_i\right\}, \quad i = 1,2,\ldots \tag{11}$$

and

$$\lim_{i\to\infty} E\left\{W'_{i+1} \mid \mathfrak{Z}_i\right\} = EW \tag{12}$$

with probability one.

Let $\sigma^2_{W'}$ denote the variance of W'.

THEOREM 1: If (11) and (12) hold and

$$\sum_{i=1}^{\infty} \frac{\sigma^2_{W'_i}}{i^2} < \infty, \tag{13}$$

then (10) holds with probability one.

PROOF: The theorem is an immediate consequence of a strong law of large numbers for dependent random variables (see [5], p. 387).

In order to apply the theorem to our replacement problem, we state the rule R'_0: Let X_{kj} denote the j^{th} observation on the amount of deterioration in the k^{th} replacement cycle. Then, let N_1 be prescribed by any arbitrary rule ($N_1 = 1$, say); $F_1(x)$ be the empirical distribution function estimate from $X_{11},\ldots,X_{1N_1}$; N_2 be the smallest n such that

$$S_{2n} = \sum_{j:1}^{n} X_{2j} \geq a_{2n} = \max\left[L - F_1^{-1}\left(1 - \frac{c}{An}\right), 0\right];$$

$F_2(x)$ be the empirical distribution function estimated from $X_{11},\ldots,X_{2N_2}$;..., etc. R'_0 is thus recursively defined.

We ask, now, whether (10) holds, where $W'_i = C_i/N_i$ and $H_i = F_i$ for $i = 1,2,\ldots$. To show that (10) does hold, we verify that (11), (12), and (13) are satisfied. We omit the measure theoretic details of the construction of the sequence $\{\mathfrak{Z}_i\}$. Briefly, $\mathfrak{Z}_i$, for $i = 1,2,\ldots$, is generated by conditions on all X_{kj} for $k \leq i$. It is clear from the nature of the rule (i.e., C_{i+1} and N_{i+1} depend on previous cycles only by way of F_i) that (11) holds. Since C_i is bounded by c + A, (13) holds. It remains only to show that (12) is satisfied. To this end, after noting that

$$C_i = c \quad \text{if } a_{in} \leq S_{in} \leq L \text{ for some } n$$

$$= c + A$$

otherwise, we can write

$$E\left\{\frac{C_{i+1}}{N_{i+1}} \,\middle|\, F_i\right\} = c\sum_{k=1}^{\infty}\frac{1}{k}P\left(N_{i+1} = k \mid F_i\right) \tag{14}$$

$$+ A\sum_{k=1}^{\infty}\frac{1}{k}P\left(N_{i+1} = k,\ S_{i+1,k} > L \mid F_i\right).$$

Now

$$P\left(N_{i+1} = k \mid F_i\right) = P\left(S_j < a_{ij},\ i \le j \le k-1,\ S_k \ge a_{ik}\right),$$

with a similar expression for $P\left(N_{i+1} = k,\ S_{i+1,k} > L \mid F_i\right)$.

If F is either continuous and strictly increasing or purely discrete, since $\lim_{i\to\infty} F_i = F$ with probability one and $0 \le a_{in} \le L$, each term of (14) converges, as $i\to\infty$, with probability one, to the same term with a_{in} replaced by a_n. A convergent dominating series can be obtained by letting $a_{in} = L$. Hence, we can interchange limits and (12) is satisfied. We have proved

THEOREM 2: If F is either continuous and strictly increasing or purely discrete, R'_0 is defined as above. $W'_i = C_i/N_i$ for $i = 1, 2, \ldots$, then (10) holds with probability one.

We feel safe in conjecturing that R^* is of the form: Replace when $S_n > \min(L, L^*)$, where L^* depends on F. If $R^{*\prime}$ is defined in a way analogous to the definition of R'_0, it is possible to prove, as in Theorem 2, that

$$\lim_{T\to\infty}\frac{\sum_{i=1}^{M(T)} C_i}{T} = \phi_{R^*}$$

with probability one.

REFERENCES

[1] Arrow, K. J., Karlin, S. and Scarf, H., Studies in the Mathematical Theory of Inventory and Production, Stanford University Press, Stanford, Calif., 1958.

[2] Doob, J., Stochastic Processes, John Wiley and Sons, Inc., New York, 1953.

[3] Johns, M. V. Jr., "Non-Parametric Empirical Bayes Procedures" Annals of Mathematical Statistics, Vol. 28, 1957, pp. 649-670.

[4] Klein, M. and Rosenberg, L., "The Deterioration of Inventory and Equipment," Naval Research Logistics Quarterly (to appear).

[5] Loève, M., Probability Theory, D. Van Nostrand, New York, 1955.

[6] Robbins, H., "An Empirical Bayes Approach to Statistics," Proceedings of the Third Berkeley Symposium on Mathematical Statistics, University of California Press, 1956, pp. 157-163.

* * *

OPTIMAL SYSTEM SUPPLY*

Frank Proschan

Electronic Defense Laboratories
Mountain View, California

This paper unifies and extends previous work by Guy Black and Frank Proschan which shows how to calculate minimum-cost spare-part kits yielding specified assurance against shutdown due to shortage in a complex military system. Optimal policies are determined for both the situation in which only the initial spare-parts kit may be used to replace failed components (single-period model) and the situation in which periodic replenishment of the spare-part kit occurs (multiperiod model). In the multiperiod model, constant lag in delivery is permitted. The only input data required are component costs and failure distributions (assumed exponential in the multiperiod model); knowledge of component usage rates is not required. Actual applications of the single-period model to certain large military systems are described briefly.

1. INTRODUCTION AND SUMMARY

In [3] Black and Proschan show how to determine spare-part kits for a complex military system (such as a radar or a computer) which, for the money expended on spare parts, provide maximum assurance against system shutdown due to shortage of essential components. The only input data needed are unit costs and component failure rates. Knowledge of the rates of usage of components is not required, so that the method is especially attractive for new systems. It is assumed in the model that the length of component life is a random variable governed by an exponential distribution, that components (which could be black boxes or subsystems) are connected in series, and that only initially provided spares could be used to replace failures.

The question immediately arose: Does the marginal analysis technique for computing optimal kits used in [3] also yield optimal kits when component-failure distributions are not exponential, but say normal, Gamma, or any of the other common statistical distributions? In [23] it is shown that for a wide class of distributions (known as monotone likelihood distributions), which includes the exponential, normal, Gamma, and other well-known distributions, the marginal analysis technique of [3] would yield optimal kits. In the course of deriving this result, a number of new results were obtained in the branch of probability known as renewal theory. These results and their extensions appear in [17].

The next logical extension of the model was to cover the periodic restocking of the spares kit. In [22] there was derived the optimal policy for the case of exponential failure distributions, with delivery being either immediate or subject to specified lag—optimal in the sense of achieving a specified protection each period against system shutdown due to shortage—at a minimum expected total cost.

* Prepared for the U.S. Army Signal Research and Development Laboratory under Signal Corps Contract DA 36-039 SC-78281 and published previously as Electronic Defense Laboratories Engineering Report EDL-E38.

The mathematical technique for computing optimal spare-part kits in the case of exponential failure distributions has been used at the Electronic Defense Laboratory to compute spares kits for three major systems [10]. In addition, the technique has been used to compute spares kits for (a) the Sylvania's Data Systems Operations subcontract for the Ballistic Missile Early Warning System, (b) the University of Washington contract for the development of a naval range finder, and (c) the Sperry-Gyroscope satellite contract.

In view of the interest in, and response to, the separate reports covering the various phases of the subject, it was felt wise to summarize and integrate the results reported in [2, 3, 22, and 23] and, in addition, to supply some missing elements of importance, such as budgeting for spares expenditures and computation of kits when knowledge of failure rates is uncertain. The present report is intended to achieve this purpose and should make it unnecessary for the reader to refer back to the original reports, some of which may be difficult to obtain at this time.

In our first model [2, 3, 23] we show how to determine the number of spare parts for each essential type of component in a complex system so as to provide maximum assurance against system shutdown due to shortage during a specified period of time, under a budget constraint. The only input data needed are unit costs and failure distributions. Two equivalent procedures are presented for computing the composition of the optimal spare-parts kit in terms of $Q_i(n)$, the probability of n or less failures of component type i, $i = 1, 2, \ldots, k$, when the $Q_i(n)$ have the monotone-likelihood ratio property. We then show how to obtain a family of optimal kits corresponding to various possible budgets available for spare parts. Next we show how to express the $Q_i(n)$ in terms of the probability densities, $f_{ij}(t)$, for component life (Sec. 2.1, 2.2).

Since the procedures for computing optimal kits are valid when the $Q_i(n)$ have the monotone likelihood ratio property, we next investigate the requirements on the $f_{ij}(t)$ which will insure that the $Q_i(n)$ have the monotone-likelihood property. By using the mathematical results of the Appendix, we show that a sufficient condition is that the $f_{ij}(t)$ have the monotone-likelihood ratio property. Since this class of probability densities includes the exponential, the normal, the Gamma, and most of the commonly occurring densities, we have a solution of fairly general applicability (Sec. 2.3.).

We next extend the model to cover the case in which successive replacements have monotone-likelihood ratio densities which are not necessarily identical. This case would arise if any of the operating components or replacements are not new. Another extension of the model treats the case in which the scheduled length of operation of a component (and its replacements) is a random variable having a monotone-likelihood ratio density; we confine ourselves to the case in which the system contains just one operating component of each type. We show that Procedures 1 and 2 of Section 2.2.1 yield optimal spare-part kits under both these extensions.

An interesting application of the model in the field of reliability is then described. In the design of a complex system to achieve maximum reliability under a weight constraint, the very same mathematical problem arises as discussed above. All that we need do is substitute "redundant standby unit" for "spare part," "weight" for "cost," and "reliability" for "assurance against system shutdown due to shortage"; and we must further assume that switching-in of a standby unit occurs with perfect reliability and that standby units do not deteriorate when not in use (Sec. 2.4.).

Next we show in detail how to compute optimal spare-part kits in the important special case of exponential failure densities, commonly assumed to characterize electron tubes. The computation is particularly simple, involving Poisson densities which are well tabulated. First we present a small-scale hypothetical example to illustrate the steps involved in the computation. Next we report several actual applications of the model to the computation of optimal spare-part kits for some rather large military systems developed by the Electronic Defense Laboratory [10]. Applications of the model at other laboratories are also described. Last, in this section we show how to obtain optimal spare-part kits when the failure rates of the exponential distribution are not known with certainty, but rather distributed according to a Gamma distribution (Sec. 2.5.).

We next consider the important class of inventory problems in which the spare-part kit is replenished at the beginning of each of a number of periods of time. We determine that resupply policy which achieves a specified protection during each period against system shutdown due to shortage, at minimum total cost (future costs are discounted to their present value, and exponential failure densities are assumed). The optimal policy turns out to be surprisingly simple and relatively easy to administer.

In the case of immediate delivery, the optimal policy specifies: (1) For the first period, provide the optimal spare-parts kit by use of Procedure 1 or 2 of Section 2.2.1. (2) At the beginning of each period, starting with the second, order enough spares to replace spares actually used during the preceding period.

In the more realistic case of lag in delivery, the optimal policy is still fairly simple and makes use of optimal spare-part kits, as computed by Procedures 1 and 2 of Section 2.2.1. Finally, an approximate expression is obtained for the amount to be budgeted to cover the cost of spares for a specified length of time with specified assurance.

In the mathematical appendix (Sec. 4) a number of new results in the theory of totally positive kernels and Polya-type distributions [17] is derived. These results are needed in demonstrating that Procedures 1 and 2 yield optimal spare-part kits when the underlying failure densities have the monotone-likelihood ratio property. A number of other applications of these new mathematical results also are made in Reference [17]. (see Sec. 4.)

2. SPARES KIT FOR SINGLE PERIOD

In Section 2.1 we state the mathematical model considered for the single-period case. In Sections 2.2 and 2.3 we obtain the solution to the model. In Section 2.4 we extend the model, though retaining the single period. Finally, in Section 2.5 we solve explicitly the special case in which the failure densities are exponential. Examples of actual applications are given.

2.1 Mathematical Model

A complex system is to be placed in the field for a period of time t_0. When a component fails, it is instantly replaced by a spare, if one is available. The components considered operate independently (i.e., failure of one does not influence failure of any other) and are essential to continued system operation, so that a shortage of any component considered results in system shutdown. Only the spares originally provided may be used for replacements (i.e., no resupply of spares can occur during the period).

The system contains d_i components of type i, i = 1, 2, ..., k, operating simultaneously. Different components of a given type may be used at different levels of intensity, so that the length of life of the j^{th} component of type i (component i, j) and its replacements are independent random variables with density $f_{ij}(t)$, j = i, ..., d_i; i = 1, ..., k. The scheduled length of time t_{ij} that component i, j and its replacements must operate during t_o is a constant, j = 1, ..., d_i; i = 1, ..., k. Finally, the cost of a single unit of type i is c_i, i = 1, ..., k.

What choice of n_i, the number of spares of type i initially provided, i = 1, ..., k, will yield maximum assurance of continued system operation (i.e., no shortage of any essential component) during the period $[0, t_o]$ at a total cost for spares $c(n_1, \ldots, n_k) = \sum_{i=1}^{k} n_i c_i \leq c_o$, a constant [3], [23]?

Related models in the spare-parts problem have been treated in [6], [7], and [8]. In those models, the expected total of weighted shortages is minimized subject to a linear weight or cost restraint, with the demand for spares described a priori by a probability density function. In this model we maximize assurance of continued system operation by optimal allocation of spares likewise subject to a linear restraint, but with the demand for spares, instead of being assumed a priori, generated by failure of operating units following known probability distributions. Thus, to obtain the composition of the optimal spare-parts kit we use information about component failure rates rather than information about component demand distributions. In the typical situation under consideration—a new system under experimentation for a single period in the field—we are thus given the opportunity to use information we may have (component failure rates) rather than called upon to provide information we may not have (component demand distributions).

Our choice of assurance of continued system operation as the figure of merit to be maximized is especially relevant in military applications, where a penalty cost is often difficult to determine. In the system reliability model of Section 2.4.3, especially, probability of successful system operation is a completely natural choice for the figure of merit.

2.2 Maximizing a Nonlinear Function Subject to a Linear Constraint

2.2.1 Maximizing $Q(\underline{n})$ as a Function of the $Q_i(n)$.

Let

$Q_i(n)$ = probability that n or less failures of type i occur during $(0, t_o)$, i = 1, ..., k,

$\underline{n} = n_1, n_2, \ldots, n_k$, the vector specifying the number of spares of each type,

$Q(\underline{n})$ = probability that no shortage is experienced for any of the k types during $(0, t_o)$, given an initial kit of composition $\underline{n}$,

$c(\underline{n})$ = total cost of a spare-parts kit composed of n_i units of type i, i = 1, 2, ..., k.

Since components operate independently and failure of any one causes system shutdown, the probability of continued system operation during $(0, t_o)$ is given by

$$Q(\underline{n}) = \prod_{i=1}^{k} Q_i(n_i). \tag{1}$$

Also, we note that

$$c(\underline{n}) = \sum_{i=1}^{k} n_i c_i .$$

We wish to maximize $Q(\underline{n})$ subject to

$$\sum_{i=1}^{k} n_i c_i \leq c_o \text{ and } n_i \geq 0,\ i = 1, \ldots, k. \tag{3}$$

Define $R_i(n) = \ln Q_i(n)$ and $R(\underline{n}) = \ln Q(\underline{n})$. Then it is equivalent to maximize $R(\underline{n})$ subject to Equation (3).

Maximizing a nonlinear funtion $R(\underline{n})$ subject to linear restraints (3) is a special case of nonlinear programming, treated in [19], where the theorems are developed in detail for continuous variables. In our problem we are dealing with discrete variables, $n_1, \ldots, n_k$. Thus we shall independently derive the required theorems.

Define $\Delta R_i(n) = R_i(n+1) - R_i(n)$ for $n = 0, 1, 2, \ldots$; $i = 1, 2, \ldots, k$. Then we shall show in Theorem 1 and its Corollary below that when $\Delta R_i(n)$ is decreasing in n for $i = 1, \ldots, k$ we may obtain the solution to our inventory problem by following Procedure 1.

PROCEDURE 1: Let $\Delta R_i(n)$ be a decreasing function of n for $i = 1, \ldots, k$. For arbitrary $r > 0$, for those i such that $\Delta R_i(0) < rc_i$, define $n_i^*(r) = 0$; for the remaining i, define $n_i^*(r)$ as 1 + largest n such that $\Delta R_i(n) \geq rc_i$. Compute $c[\underline{n}^*(r)] = \sum_{i=1}^{k} c_i n_i^*(r)$. Let r_o be the value of r yielding maximum $c[\underline{n}^*(r)] \leq c_o$. Then $\underline{n}^*(r_o)$ is the spares kit corresponding to a budget restraint of c_o.

The following theorem shows $\underline{n}^*$ is optimal when c_o is one of the values assumed by $c[\underline{n}^*(r)]$ as r varies over $(0, \infty)$:

THEOREM 1: $\underline{n}^*$ maximizes $R(\underline{n})$ among all $\underline{n}$ such that $c(n) \leq c(\underline{n}^*)$, $\underline{n} \geq 0$.

PROOF: We will show for any $\underline{0} \leq \underline{n} \neq \underline{n}^*$ for which $c(\underline{n}) \leq c(\underline{n}^*)$ that $R(\underline{n}) \leq R(\underline{n}^*)$. Suppose $n_i > n_i^*$ for i in I_1, $n_i < n_i^*$ for i in I_2, where I_1, I_2 are subsets of $\{1, 2, \ldots, k\}$. Since $\Delta R_i(n)$ is a decreasing function of n by assumption, then

$$\Delta R_i(n_i^* + j) \leq rc_i \text{ for } i \text{ in } I_1,\ j = 0, 1, 2, \ldots, n_i - n_i^* - 1. \tag{4a}$$

Similarly, for i in I_2,

$$\Delta R_i(n_i^* - j) \geq rc_i \text{ for } i \text{ in } I_2,\ j = 1, 2, \ldots, n_i^* - n_i. \tag{4b}$$

Hence

$$R(\underline{n}) - R(\underline{n}^*) = \sum_{i \text{ in } I_1} \sum_{j=0}^{n_i - n_i^* - 1} \Delta R_i(n_i^* + j) - \sum_{i \text{ in } I_2} \sum_{j=1}^{n_i^* - n_i} \Delta R_i(n_i^* - j)$$

$$\leq r \sum_{i \text{ in } I_1} (n_i - n_i^*)c_i - r \sum_{i \text{ in } I_2} (n_i^* - n_i)c_i = r \sum_{i=1}^{k} (n_i - n_i^*)c_i$$

$$= r[c(\underline{n}) - c(\underline{n}^*)].$$

But $r > 0$ and $c(\underline{n}) - c(\underline{n}^*) \leq 0$. Hence $R(\underline{n}) \leq R(\underline{n}^*)$.

QED

(In the remainder of this section, we shall confine attention to the case in which $\Delta R_i(n)$ is a decreasing function of n for $i = 1, \ldots, k$. In Section 2.3, we derive a sufficient condition for $\Delta R_i(n)$ decreasing.)

It follows that a given value of r yields a kit composition $\underline{n}^*(r)$ which solves the original inventory problem for the particular case $c_0 = \sum_{i=1}^{k} c_i n_i^*(r)$. Hence, by varying r over the interval $(0, \infty)$, we may obtain solutions corresponding to a c_0 equal to any of the different (discrete) values $\sum_{i=1}^{k} c_i n_i^*(r)$ may assume. We shall refer to any vector $\underline{n}^*$ obtained by Procedure 1 as an optimal point or an optimal kit.

For values of c_0 different from any of the $c(n^*)$, Procedure 1 may provide only an approximation solution because of the discrete nature of the variables $n_1, \ldots, n_k$; the bound on the error is given in the Corollary below and in most practical problems will be small.

COROLLARY: For $c_0 \neq c[\underline{n}^*(r)]$ for all $r > 0$,

$$\max_{\substack{c(\underline{n}) \leq c_0 \\ \underline{n} \geq 0}} Q(\underline{n}) - Q(\underline{n}^*(r_0)) \leq Q(\underline{n}') - Q[\underline{n}^*(r_0)]$$

where $c(\underline{n}') = \min_{c(\underline{n}^*) > c_0} c(\underline{n}^*)$. ($\underline{n}'$ represents the cheapest optimal kit of higher cost than c_0.)

PROOF: By Theorem 1, $R(\underline{n}') \geq R(\underline{n})$ for all $\underline{n} \geq 0$ such that $c(\underline{n}) \leq c_0 < c(\underline{n}')$. Hence $Q(\underline{n}') \geq Q(\underline{n})$ for all $\underline{n} \geq 0$ such that $c(\underline{n}) \leq c_0 < c(\underline{n}')$. Thus

$$Q(\underline{n}') \geq \max_{\substack{c(\underline{n}) \leq c_0 \\ \underline{n} \geq 0}} Q(\underline{n}).$$

Hence the theorem.

QED

Thus the error in using Procedure 1 is, at most, the difference in protection achieved by the two adjacent optimal kits whose costs straddle the specified c_0. This motivates the following alternate procedure for obtaining in succession the same set of optimal points as are obtained by Procedure 1.

PROCEDURE 2: First note that $(0, \ldots, 0)$ is an optimal point (corresponding to $c_0 = 0$). Next let $\underline{n}^*$ be any optimal point. Compute $\Delta R_i(n_i^*)/c_i$ for $i = 1, \ldots, k$. If the α^{th} ratio is the largest, the next optimal point to the right of $\underline{n}^*$ is $(n_1^*, \ldots, n_{\alpha-1}^*, n_\alpha^* + 1, n_{\alpha+1}^*, \ldots, n_k^*)$, as will be proved in Theorem 2. (We shall designate the immediate successor of an optimal point $\underline{n}^*$ by $\underline{n}^*(\alpha)$, where α is the (only) coordinate that has changed.)

THEOREM 2: Any point obtainable by Procedure 1 is obtainable by Procedure 2, and conversely.

PROOF: Let $\underline{n}^*$ be a point obtainable by Procedure 2. Suppose $\underline{n}^*$ is obtained from the preceding point on the left by a change in the α^{th} coordinate. Set $r_0 = [\Delta R_\alpha(n_\alpha^* - 1)]/c_\alpha$.

Thus

$$\frac{\Delta R_\alpha^*(n_\alpha^*)}{c_\alpha} < r_o \leqq \frac{\Delta R_\alpha(n_\alpha^* - 1)}{c_\alpha}, \tag{5}$$

since $\Delta R_\alpha(n)$ is decreasing. Also we know

$$\frac{\Delta R_i(n_i^*)}{c_i} < r_o \tag{6}$$

for all $i \neq \alpha$. Now if there existed β such that $[\Delta R_\beta(n_\beta^* - 1)]/c_\beta < r_o \leqq [\Delta R_\alpha(n_\alpha^* - 1)]/c_\alpha$, there would be a point lying to the left of $\underline{n}^*$ whose α^{th} coordinate would be n_α^* and whose β^{th} coordinate would be $\leqq n_\beta^* - 1$ (i.e., in computing successive points, the α^{th} coordinate would change before the β^{th}). This would contradict the assumption that $\underline{n}^*$ is obtained from the preceding point on the left by a change in the α^{th} coordinate. Thus we conclude that

$$\frac{\Delta R_i(n_i^* - 1)}{c_i} \geqq r_o \tag{7}$$

for all $i \neq \alpha$. From (5), (6), and (7), we see that $\underline{n}^*$ is obtainable by Procedure 1.

Next, let $\underline{n}^*$ and $\underline{n}^*(\alpha)$ be a pair of successive points obtained by use of Procedure 2. We shall show that there are no points lying between them obtainable by Procedure 1. Assume there is such a point $\underline{m}^*$ with $r = r_o$. For at least one coordinate of $\underline{m}^*$ other than the α^{th}, say the β^{th}, we thus have either (1) $m_\beta^* > n_\beta^*$ or (2) $m_\beta^* \leqq n_\beta^*$. Assume (1). Then the β^{th} coordinate of $\underline{n}^*$, $\underline{m}^*$, and $\underline{n}^*(\alpha)$ are, respectively, n_β^*, m_β^*, and n_β^*. Since $\underline{n}^*$ and $\underline{n}^*(\alpha)$ are obtainable by Procedure 2, we have just shown that each is obtainable by Procedure 1. Hence there is an $r = r_1$ and $r = r_2$, with $r_1 > r_2$ associated with $\underline{n}^*$ and $\underline{n}^*(\alpha)$, respectively, with the property that $[\Delta R_\beta(n_\beta^* - 1)]/c_\beta \geq r_1$ and $[\Delta R_\beta(n_\beta^*)]/c_\beta < r_2$. Also $[\Delta R_\beta(m_\beta^* - 1)]/c_\beta \geqq r_o$, while $[\Delta R_\beta(m_\beta^*)]/c_\beta < r_o$. But $\underline{m}^*$ lies to the left of $\underline{n}^*(\alpha)$, so that $r_o > r_2$; hence we have $[\Delta R_\beta(n_\beta^*)]/c_\beta < r_2 < r_o \leqq [\Delta R_\beta(m_\beta^* - 1)]/c_\beta$. This is impossible since $n_\beta^* \leqq m_\beta^* - 1$. A similar argument rules out (2). Thus there are no points obtainable by Procedure 1 lying between $\underline{n}^*$ and $\underline{n}^*(\alpha)$. This means that every point obtainable by Procedure 1 is obtainable by Procedure 2.

QED

Procedure 2 represents a systematic way of obtaining all optimal points needed and is more convenient for machine computation. Actually the two methods may be usefully combined in the following commonly occurring situation: Suppose there is no single value of c_o specified. Rather, suppose what is desired is a curve (optimal curve) showing maximum protection $Q(\underline{n}^*)$ against system shutdown as a function of total cost $c(\underline{n}^*)$ for spares. To obtain this curve, compute an optimal point by Procedure 1 near the lower end of the range of interest of protection or cost (several trial values of r may have to be used). Then proceed systematically up the curve by Procedure 2. The following theorem furnishes some helpful information about the number of points in any portion of the optimal curve:

THEOREM 3: The number of points on the optimal curve between any two distinct optimal points $\underline{m}^*$ and $\underline{n}^*$ (with $\underline{m}^* \leq \underline{n}^*$, say) is $\sum_{i=1}^{k} (n_i^* - m_i^*) - 1$.

PROOF: By Procedure 2 we get successive points by changing one coordinate at a time. The number of such changes in going from $\underline{m}^*$ to $\underline{n}^*$ is $\sum_{i=1}^{k} (n_i^* - m_i^*)$. Excluding the last change which results in $\underline{n}^*$, we find that there are $\sum_{i=1}^{k} (n_i^* - m_i^*) - 1$ optimal points between $\underline{m}^*$ and $\underline{n}^*$.

QED

2.2.2 Expressing $Q_i(n)$ as a Function of the f_{ij}. Thus far, we have shown how the problem may be solved in terms of $Q_i(n)$, $i = 1, \ldots, k$, under the assumption $Q_i(n+1)/Q_i(n)$ is decreasing for each i. In the remainder of this section we shall obtain the relationship between the $Q_i(n)$ and the underlying failure densities $f_{ij}(t)$. Clearly the random process describing successive replacement of failed components is a renewal process; thus the continued operation of the complex system may be thought of as a set of independent renewal processes.

For $j = 1, \ldots, d_i$; $i = 1, \ldots, k$; $-\infty < t < \infty$, let

$$p_{ij}(1,t) = f_{ij}(t)$$

$$p_{ij}(2,t) = \int_{-\infty}^{\infty} f_{ij}(\theta)\, f_{ij}(t-\theta)\, d\theta$$

$\ldots$

$$p_{ij}(n,t) = \int_{-\infty}^{\infty} f_{ij}(\theta)\, p_{ij}(n-1, t-\theta)\, d\theta \qquad \text{for } n = 3, 4, \ldots$$

$$P_{ij}(n,t) = \int_{t}^{\infty} p_{ij}(n,\theta)\, d\theta \qquad \text{for } n = 1, 2, \ldots$$

$$\Delta P_{ij}(n,t) = \begin{cases} P_{ij}(1,t) & \text{for } n = 0 \\ P_{ij}(n+1,t) - P_{ij}(n,t) & \text{for } n = 1, 2, \ldots \end{cases}$$

$q_i(n)$ = probability of exactly n failures of component type i.

Thus $p_{ij}(n,t)$, being the n-fold convolution of $f_{ij}(t)$, represents the probability density for the combined life of the original i, j component and n-1 replacements; as a consequence, $P_{ij}(n, t_{ij})$ is the probability that n-1 or less spares will be used in replacing the i, j component during the specified t_{ij} units of time. From the definition of $\Delta P_{ij}(n, t_{ij})$, we know from renewal theory [1, p. 272] that $\Delta P_{ij}(n, t_{ij})$ represents the probability that exactly n spares will be required during the t_{ij} units of time scheduled for component i, j (and its replacements). Next, since the total number of spares of type i required for continued system operation during $(0, t_o)$ is simply the sum of the numbers required to replace component i, j for $j = 1, 2, \ldots, d_i$, we have

$$q_i(n) = \Delta P_{i1}(n, t_{i1}) * \Delta P_{i2}(n, t_{i2}) \cdots * \Delta P_{id_i}(n, t_{id_i}),$$

where the convolution (indicated by *) is taken in the first argument of the $\Delta P_{ij}(n, t_{ij})$. Finally, $Q_i(n)$, the probability of n or less failures of type i during $(0, t_o)$, is clearly given by

$$Q_i(n) = \sum_{m=0}^{n} q_i(m), \tag{2}$$

so that $q_i(n)$ represents the density and $Q_i(n)$ the distribution function for the total number of failures (or equivalently, the total number of spares required) of type i during $(0, t_o)$.

Thus, following the definitions above and (1) and (2), we may compute according to standard procedures (possibly somewhat tediously) the values of $Q_i(n)$ as a function of the f_{ij}. From the values of $Q_i(n)$, using either Procedure 1 or Procedure 2 (or a combination), we may solve the original inventory problem provided $[Q_i(n + 1)]/[Q_i(n)]$ is decreasing. In the next section, we shall demonstrate that $[Q_i(n + 1)]/[Q_i(n)]$ decreases whenever the $f_{ij}(t)$ are monotone likelihood ratio [1] densities in differences of t.

2.3 Sufficient Condition for Use of Procedures 1 and 2

THEOREM 1: If the underlying failure densities f_{ij}, $j = 1, 2, \ldots, d$ and $i = 1, 2, \ldots, k$ have the monotone likelihood ratio property, then Procedures 1 and 2 of Section 2.2.1 yield optimal spare-part kits. That is, the hypothesis of Theorem 1 of Section 2.2.1 holds, so that the conclusion and Corollary follow.

PROOF:

1. $\Delta P_{ij}(n, t)$, the probability of requiring n replacements of operating component i, j, has the monotone likelihood ratio property in n for each fixed i, j by Theorem 3 of Section 4.4.

2. $q_i(n)$, the probability of requiring n replacements of type i, has the monotone likelihood ratio property in n for each i by Lemma 3 of Section 4.1, since $q_i(n)$ is obtained as the convolution of ΔP_{i1}, ΔP_{i2}, $\ldots$, ΔP_{id_i}. (See Equation (1) of the last section.)

3. $Q_i(n)$ has the monotone likelihood ratio property in n for each i, since

   (a) $Q_i(n) = \sum_{m=-\infty}^{\infty} q_i(n-m)\, H(m)$, where $H(m) = \begin{cases} 1 \text{ for } m = 0, 1, 2, \ldots \\ 0 \text{ otherwise,} \end{cases}$

   (b) H(m) has the monotone likelihood ratio property [25, p. 335],

   (c) The convolution of monotone likelihood ratio sequences has the monotone-likelihood ratio property, by Lemma 3 of Section 4.1.

4. For each i, $Q_i(n + 1)/Q_i(n)$ is decreasing as a function of n, since from statement 3 just above, for each i

[1] A probability density f(x) has the monotone likelihood ratio property if, given $x_1 < x_2$, $t_1 < t_2$, then

$$\begin{vmatrix} f(x_1 - t_1) & f(x_1 - t_2) \\ f(x_2 - t_1) & f(x_2 - t_2) \end{vmatrix} \geqq 0.$$

Further discussion and extension of this concept is given in the Appendix. Note for future reference that a monotone likelihood ratio density is a Polya frequency function of order 2 (PF_2), and conversely. (See the Appendix for a definition and discussion of Polya frequency functions.)

$$\frac{Q_i(n+1)}{Q_i(n)} - \frac{Q_i(n)}{Q_i(n-1)} = \begin{vmatrix} Q_i(n+1) & Q_i(n) \\ Q_i(n) & Q_i(n-1) \end{vmatrix} \leq 0$$

for $n = 1, 2, \ldots$.

5. Consequently, $\Delta R_i(n) = R_i(n+1) - R_i(n) = \ln \left\{ Q_i(n+1)/Q_i(n) \right\}$ is decreasing in n. Hence Procedures 1 and 2 yield optimal kits by Theorems 1 and 2 of Section 2.2.1.

2.4 Extensions of the Model

2.4.1 Nonidentical Failure Densities. It is clear from the theorems of the Appendix that the successive convolutions yielding $p(n,t)$ and $P(n,t)$ need not necessarily refer to a common PF density. Thus we may conclude that Procedures 1 and 2 of Section 2.2.1 for obtaining the optimal spares kits apply even when the successive replacements of a given component have different (known) monotone likelihood ratio failure distributions. At first glance, this may seem like a generalization of little application, since in most cases successive replacements of any one component will be made with units having a common failure distribution. However, if at the beginning of the period the components already in the system have been aged by previous use, the (conditional) distribution of remaining life is different from the distribution of life of future replacements (except only in the case of exponential life distributions). Another possibility is that replacements may be of different known ages. In this connection, it is helpful to note that the conditional probability density for the remaining life t of a component, given its age t_o, is PF_2 if the underlying density is PF_2; i.e.:

THEOREM 1: If $f(t)$ is PF_2 with $f(t) = 0$ for $t < 0$, then

$$g(t) = \begin{cases} 0 & \text{for } t < 0 \\ \dfrac{f(t_o+t)}{\int_{t_o}^{\infty} f(u)\, du} & \text{for } t \geq 0 \end{cases} \quad \text{is } PF_2.$$

PROOF: Let $t_1 < t_2$, $\theta_1 < \theta_2$. Define

$$D = \begin{vmatrix} g(t_1-\theta_1) & g(t_1-\theta_2) \\ g(t_2-\theta_2) & g(t_2-\theta_2) \end{vmatrix},$$

$$D' = \frac{1}{\int_{t_o}^{\infty} f(u)\, du} \begin{vmatrix} f(t_o+t_1-\theta_1) & f(t_o+t_1-\theta_2) \\ f(t_o+t_2-\theta_1) & f(t_o+t_2-\theta_2) \end{vmatrix}.$$

Then if $t_1 \geq \theta_2$, $D = D'$. Thus $\operatorname{sgn} D = \operatorname{sgn} D' \geq 0$ since f is PF_2. If $t_1 < \theta_2$, then $g(t_1-\theta_2) = 0$ by definition, so that $D \geq 0$.

QED

2.4.2 Scheduled Times Random. Suppose the t_{ij} are not constants but positive random variables T_{ij} with corresponding PF_2 densities $g_{ij}(t)$, $j = 1, \ldots, d_i$ and $i = 1, \ldots, k$. We shall show that when $d_i = 1$, $i = 1, \ldots, k$, (only one operating component of each type) the probability density for the number of replacements of component i during $(0, t_i)$ is PF_2, so that we may conclude that Procedures 1 and 2 of Section 2.2.1 yield optimal spare-parts kits, as before.

We shall drop the subscript i in what follows. Define

$$\phi(n,\omega) = \begin{cases} 0 & \text{for } n < 0 \\ \text{probability that the sum of } n+1 \text{ independent observations from } f(t) \text{ exceeds } T \text{ by at least } \omega, \text{ where } -\infty < \omega < \infty, \text{ for } n = 0, 1, 2, \ldots . \end{cases}$$

Then

$$\phi(n,\omega) = \int P(n+1,t)\, g(t-\omega)\, dt \quad \text{for } n = 0, 1, 2, \ldots .$$

Thus, since $P(n,t)$ is TP_2 as an easy consequence[2] of Theorem 1 of Section 4.2 and since $g(t)$ is PF_2 by assumption, we conclude that $\phi(n,\omega)$ is TP_2 (where n ranges over 0, 1, 2, ...) by Lemma 2 of Section 4.1.

Next we shall show $\phi(n-m,\omega)$ is TP_2 in n and m for each fixed real ω. Let $n_1 < n_2$, $m_1 < m_2$. Define

$$D = \begin{vmatrix} \phi(n_1-m_1,\omega) & \phi(n_1-m_2,\omega) \\ \phi(n_2-m_1,\omega) & \phi(n_2-m_2,\omega) \end{vmatrix}$$

(a) If $n_1 < m_2$, then $\phi(n_1-m_2,\omega) = 0$ by definition. Hence $D \geq 0$.

(b) Assume $m_1 < m_2 \leq n_1 < n_2$. Thus, we may write

$$D = \int \begin{vmatrix} \phi(n_1-m_1,\omega) & \phi(n_1-m_2,\omega) \\ \phi(n_1-m_1,\omega-\theta) & \phi(n_1-m_2,\omega-\theta) \end{vmatrix} p(n_2-n_1,\theta)\, d\theta .$$

But the determinant in the integrand ≥ 0 since $\phi(n,\omega)$ is TP_2 (where n ranges over 0, 1, 2, ...). Hence $\phi(n-m,\omega)$ is TP_2 in n and m for each fixed real ω.

Thus $[\phi(n+1,\omega)]/[\phi(n,\omega)]$ decreases in n for fixed ω. But $\phi(n,0)$ represents the probability that n or less spares will be required during (0, T). With the use of similar arguments

---

[2] A function $g(x, y)$ is totally positive of order 2 (TP_2) if, given $x_1 < x_2$, $y_1 < y_2$, then $\begin{vmatrix} g(x_1,y_1) & g(x_1,y_2) \\ g(x_2,y_1) & g(x_2,y_2) \end{vmatrix} \geq 0$. (See the Appendix for a further discussion of totally positive functions.)

to those of Section 2.3, the monotonicity of the ratio $\phi(n+1,0)/\phi(n,0)$ is sufficient to insure that Procedures 1 and 2 yield optimal spare-parts kits.

2.4.3 Application to Reliability Design. It is interesting to note that the solution of the inventory model of Section 2.1 may be applied to a similar model in the field of reliability. Suppose we are designing a complex system (say a missile) and we wish to attain maximum reliability by providing redundant units within a weight restraint. The statement of the inventory problem of Section 2.1 then describes the present reliability problem if we make certain minor modifications. Simply substitute "redundant standby unit" for "spare," "weight" for "cost." We assume, too, that (a) switching-in of a redundant unit occurs with perfect reliability, and (b) standby units are not actually in use and therefore have no probability of failure during standby.

It is then clear that the inventory problem and the present reliability problem are mathematically the same. Hence Procedures 1 and 2 of Section 2.2.1 yield the required number of redundant units of each component type to provide maximum reliability under a weight restraint.

2.5 Exponential Failure Distributions

2.5.1 Explicit Solution. In the case of a complex electronic system, the component failure distribution usually assumed is the exponential. We shall now obtain an explicit solution of the inventory problem for this case and present illustrations.

Assume, then, that

$$f_{ij}(t) = \begin{cases} 0 & \text{for } t < 0 \\ \mu_{ij} e^{-\mu_{ij} t} & \text{for } t \geqq 0 \end{cases},$$

where μ_{ij} represents the failure rate per unit of time, $j = 1, \ldots, d_i$ and $i = 1, \ldots, k$. It follows immediately that $\Delta P_{ij}(n, t_{ij}) = e^{-\mu_{ij} t_{ij}} (\mu_{ij} t_{ij})^n / n!$, a Poisson density with parameter $\mu_{ij} t_{ij}$ [1, p. 272]. Next note that

$$q_i(n) = e^{-\lambda_i} \frac{\lambda_i^n}{n!} \tag{1}$$

where $\lambda_i = \sum_{j=1}^{d_i} \mu_{ij} t_{ij}$, since the convolution of Poisson densities is a Poisson density with parameter given by the sum of the separate parameters [4, p. 205].

We know that the exponential density is PF_2 [15, p. 125]. Hence Procedures 1 and 2 yield the optimal spare-parts kits by Theorem 1 of Section 2.3. In this connection, it is interesting to give a separate proof that $\Delta R_i(n)$ is a decreasing function of n for fixed t, the condition needed to apply Theorem 1 of Section 2.2.1.

THEOREM 1: If $f_{ij}(t) = \begin{cases} 0 & \text{for } t < 0 \\ \mu_{ij} e^{-\mu_{ij} t} & \text{for } t \geq 0 \end{cases}$, $j = 1, \ldots, d_i$, then $\Delta R_i(n)$ is a decreasing function of n.

PROOF: With the subscripts dropped $\Delta R(n) = \ln [Q(n+1)/Q(n)] = \ln \left\{ 1 + [\lambda^{n+1}/(n+1)!] \Big/ \sum_{j=0}^{n} [\lambda^j/j!] \right\}$. It will be sufficient to show that

$$g(n, \lambda) = \frac{\frac{\lambda^{n+1}}{(n+1)!}}{\sum_{j=0}^{n} \frac{\lambda^j}{j!}}$$

is a decreasing function of n for all $\lambda > 0$.

Now $g(n, \lambda) - g(n-1, \lambda)$ has the same sign as $f(n, \lambda) = \lambda \sum_{j=0}^{n-1} (\lambda^j/j!) - (n+1) \sum_{j=0}^{n} (\lambda^j/j!)$.

But, after simplification,

$$\frac{df(n, \lambda)}{d\lambda} = f(n-1, \lambda), \tag{2}$$

with $f(1, \lambda) = -2-\lambda < 0$ for $\lambda > 0$. Suppose $f(n, \lambda) < 0$ for $n = 1, 2, \ldots, n_0-1$. Then $df(n_0, \lambda)/d\lambda < 0$ for $\lambda > 0$ by Equation (2). Since $f(n_0, 0) = -(n_0+1) < 0$ and $df(n_0, \lambda)/d\lambda < 0$, then $f(n_0, \lambda) < 0$ for all $\lambda > 0$. By induction, $f(n, \lambda) < 0$ for $n = 1, 2, \ldots; \lambda > 0$. Thus $g(n, \lambda)$ is a decreasing function of n for $\lambda > 0$.

QED

A useful approximation in the case of exponential failure distributions may be derived as follows: Note that

$$\Delta R_i(n) = \ln \left[1 + \frac{\frac{\lambda_i^{n+1} e^{-\lambda_i}}{(n+1)!}}{\sum_{j=0}^{n} \frac{\lambda_i^j e^{-\lambda_i}}{j!}} \right]$$

For n such that $\sum_{j=0}^{n} \left[\left(\lambda_i^j e^{-\lambda_i} \right)/(j!) \right]$ is close to 1, and hence $\left[\left(\lambda_i^{n+1} e^{-\lambda_i} \right)/(n+1)! \right]$ close to 0,

we may write

$$\Delta R_i(n) \approx \frac{\frac{\lambda_i^{n+1} e^{-\lambda_i}}{(n+1)!}}{\sum_{j=0}^{n} \frac{\lambda_i^j e^{-\lambda_i}}{j!}},$$

so that

$$\Delta R_i(n) \approx \frac{\lambda_i^{n+1} e^{-\lambda_i}}{(n+1)!}. \tag{3}$$

Since the latter expression is tabulated [20], the computation under Procedure 1 or 2 is simple even with only a desk calculator.

2.5.2 Hypothetical Illustration. A system consisting of a UHF receiving subsystem and a VHF receiving subsystem is to be placed in the field for a three months period of experimentation. The essential tubes in the two subsystems are described in Table 1.

During the period of operation in the field, the UHF tubes are each scheduled for 332 hours of use and the VHF tubes are each scheduled for 2160 hours of use. By assuming both an exponential life distribution for each of the tube types with failure rate as shown above and an independence of operation of the tubes, find an optimal allocation of spare parts for various spares budgets (i.e., an allocation which maximizes assurance of continued system operation during the period in the field).

TABLE 1

i	Tube Type	Failure Rate/Hour	c_i, Cost per Tube ($)	Number in UHF, Scheduled for 332 Hours of Use Each	Number in VHF, Scheduled for 2160 Hours of Use Each	λ_i, Expected Number of Failures
1	Radechon	1/2500	240	4	4	4.0
2	Memotron	1/4000	1025	2	5	2.9
3	Carcinotron	1/800	1158	4	0	1.7
4	TWT	1/6000	750	2	0	.11

First, we compute the expected number of spares of each type used during the period:

$$\lambda_1 = 1/2500 \ \{4 \cdot 332 + 4 \cdot 2160\} = 4.0,$$
$$\lambda_2 = 1/4000 \ \{2 \cdot 332 + 5 \cdot 2160\} = 2.9,$$
$$\lambda_3 = 1/800 \ \{4 \cdot 332\} = 1.7,$$
$$\lambda_4 = 1/6000 \ \{2 \cdot 332\} = .11$$

One way to determine the first value of r to use is to compute $\lambda_1 + 3\sqrt{\lambda_1}$ and round to the nearest integer, obtaining 10. Let $n_1^* = 10$. Thus n_1^* corresponds to a value three standard deviations above the mean. (In a Poisson distribution, the standard deviation equals the square root of the mean.)

By the use of approximation (3), we let r be determined from $r = \left[\frac{1}{c_1} e^{-\lambda_1}\right] (\lambda_1^{10}/10!) = 0.000022$. (See Molina's Table I, [20].) We then find n_2^* as the largest value of n such that

$$\frac{1}{c_2} e^{-\lambda_1} \frac{\lambda_2^n}{n!} \geq 0.000022 . \tag{4}$$

By using Molina's Table I [20], we find $n_2^* = 6$.

By replacing the subscript 2 in Equation (4) by 3 and 4, respectively, and proceeding similarly, we find $n_3^* = 4$ and $n_4^* = 1$.

From $\underline{n}^* = 10, 6, 4, 1$, we compute $Q(\underline{n}^*) = \prod_{i=1}^{4} \sum_{x=0}^{n_i^*} e^{-\lambda_i} \lambda_i^x/x! = 0.935$, and

$$c(\underline{n}^*) = \sum_{i=1}^{4} c_i n_i^* = \$13,932.$$

Thus to obtain maximum assurance of continued system operation under a budget of \$13,932 for spares of the 4 tube types, we would stock 10 radechons, 6 memotrons, 4 Carcinotrons, and 1 TWT. The assurance obtained would be 0.935.

By proceeding in a similar fashion, we would obtain the other points plotted in Figure 1. Thus Figure 1 shows the maximum assurance of continued system operation obtainable for various given budgets for spares. In addition, the composition of the spares kit yielding the plotted maximum assurance is shown next to each point.

Given any value of c_0, we may read off Figure 1 the protection and composition of the optimal spare parts kit of cost not exceeding c_0. It would be necessary to compute two successive optimal points, $\underline{n}^*$ and $\underline{n}^{*(\alpha)}$, on the curve such that $c(\underline{n}^*) \leq c_0$ while $c(\underline{n}^{*(\alpha)}) > c_0$. $\underline{n}^*$ would then constitute the solution, the error being at most $Q(\underline{n}^{*(\alpha)}) - Q(\underline{n}^*)$.

2.5.3 Actual Applications. At the Electronic Defense Laboratory, optimal spares kits were computed by using Procedure 1 of Section 2.2.1 for three major systems (call them A, B, and C) for a one-year period [10]. The component types considered for the kits were parts peculiar and long lead time parts; i.e., parts whose shortage would lead to system shutdown for an intolerable length of time.

An exponential failure distribution was assumed for each component; the failure rate was derived from RCA Manual TR1100 [24], based on intensity of application. The actual

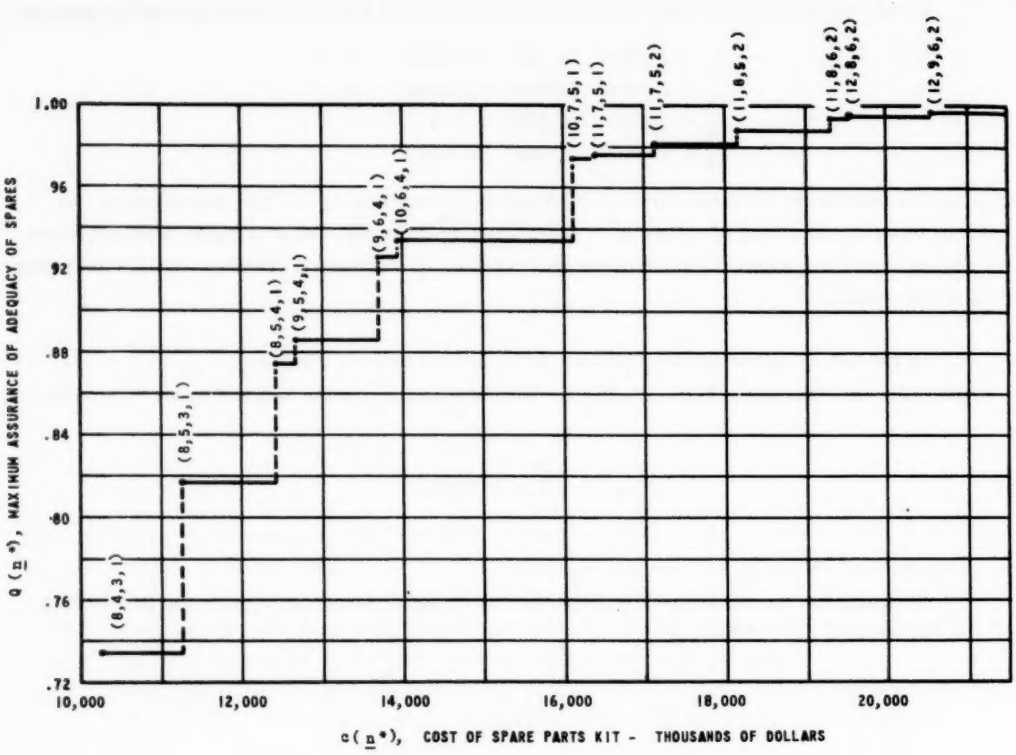


Figure 1 - Cost-assurance curve for optimal kits

computation was performed on the IBM 650, and about 25 minutes of machine time per system was required to compute the composition, assurance, and cost of 5 kits whose assurance levels ranged from 0.75 to 0.99. The results are summarized in Table 2 and Figure 2. Only the kits corresponding to an assurance between 0.85 and 0.90 are recorded in Table 2, since an assurance in this range was desired. In Figure 2, the cost-assurance curves for optimal kits for the three systems are shown. The curves are plotted as continuous, although actually they are step functions.

In programming the computation, it is convenient to determine n_i^* as the smallest integer n satisfying

$$\frac{e^{-\lambda_i}\frac{\lambda_i^{n+1}}{(n+1)!}}{\sum_{j=0}^{n} e^{-\lambda_i}\frac{\lambda_i^j}{j!}} < e^{rc_i} - 1 . \tag{1}$$

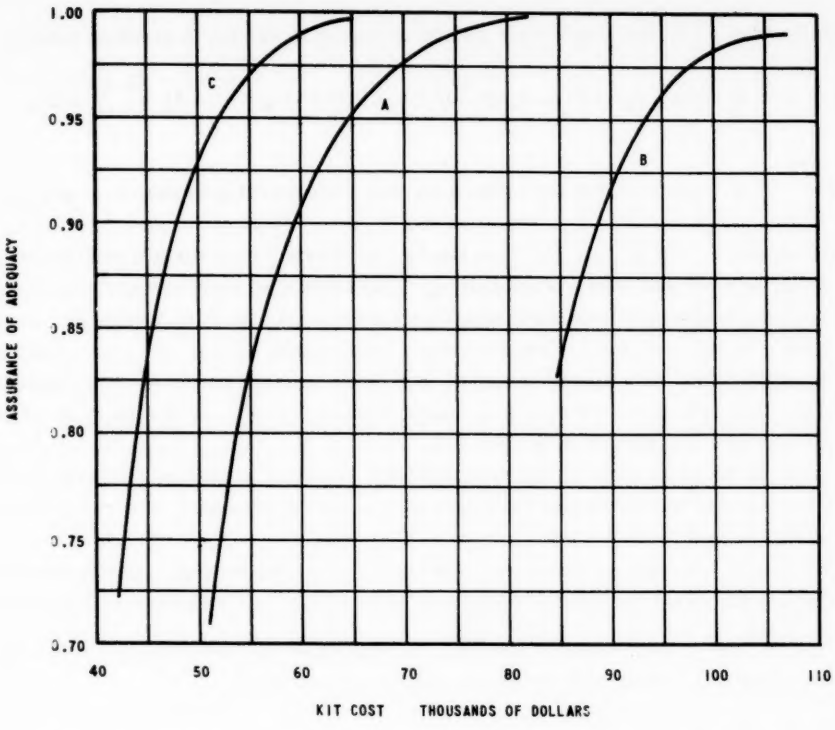


Figure 2 - Cost-assurance curves for optimal kits for three EDL systems

TABLE 2*

Optimal Kits For Three EDL Systems

System	Number of Component Types in Kit	Kit Cost	Assurance Against Shutdown Due to Shortage
A	173	$58,817	.895
B	251	87,272	.880
C	342	48,003	.895

*Extracted from Jorgensen [10, Table 2].

The retention of $e^{-\lambda_i}$ in both numerator and denominator is helpful in avoiding machine overflow. It is also desirable first to compute for fixed i (starting with i=1) n_i^*, $\sum_{j=1}^{i} n_j^* c_j$, $\prod_{j=1}^{i} \sum_{x=0}^{n_j^*} e^{-\lambda_j} \lambda_j^x / x!$ successively for each of a group of decreasing values of r and then to run through the values i = 1, 2, ..., k. This has two advantages over the alternative sequence in which i is varied first and then r is varied: (1) Kits containing a great many component types may be computed this way without exceeding the storage capacity of the computer, and (2) less computing time is required since, when we take a new, smaller value of r, we need look at possible solutions of (1) only among values of n at least as large as the n_i^* corresponding to the last r used. This reduction in effort is a consequence of the fact that the ratio in (1) is decreasing in n (See Theorem 1 of Section 2.5.1.)

2.5.4 Uncertain Knowledge of Failure Rates. In many situations, the true failure rates are not known exactly but rather are estimated with some uncertainty. How can we still provide an optimal spares kit?

Specifically, we assume the model of Section 2.5.1, but with the additional feature that λ_i, the expected number of failures of component type i, is itself a random variable with density $a_i^{b_i+1} e^{-a_i\lambda} \lambda^{b_i} / b_i!$, a member of the Gamma family. In the case, the probability $q_i(n)$ of experiencing exactly n failures of component type i is given by

$$q_i(n) = \int_0^\infty e^{-\lambda} \frac{\lambda^n}{n!} \frac{a_i^{b_i+1} e^{-a_i\lambda} \lambda^{b_i}}{b_i!} d\lambda$$

$$= \frac{a_i^{b_i+1}}{b_i!\, n!} \int_0^\infty e^{-(1+a_i)\lambda} \lambda^{n+b_i} d\lambda$$

$$= \frac{a_i^{b_i+1} (n+b_i)!}{b_i!\, n!\, (1+a_i)^{n+b_i+1}} = \left(\frac{a_i}{1+a_i}\right)^{b_i+1} \binom{n+b_i}{b_i} \frac{1}{(1+a_i)^n}.$$

Thus the density $q_i(n)$ is a negative binomial density.

In order to apply Procedure 1 or Procedure 2 to obtain optimal kits, we must verify that each $\Delta R_i(n)$ is a decreasing function of n. This, in turn, will follow if we can show that each $q_i(n)$ has the monotone likelihood ratio property in differences of n. (See statements 2, 3, and 4 in the Proof of Theorem 1 of Section 2.3.) For convenience, we drop the subscripts in proving

THEOREM 1: $$q(n) = \begin{cases} \left(\frac{a}{1+a}\right)^{b+1} \binom{n+b}{b} \frac{1}{(1+a)^n} & \text{for } n \geq 0 \\ 0 & \text{for } n < 0 \end{cases}$$

has the monotone likelihood property in differences of n.

PROOF: Let $n_1 < n_2$, $m_1 < m_2$. We must prove

$$D = \begin{vmatrix} q(n_1-m_1) & q(n_1-m_2) \\ q(n_2-m_1) & q(n_2-m_2) \end{vmatrix} \geq 0.$$

If $n_1 < m_2$, then $q(n_1-m_2) = 0$, so that the desired conclusion follows. Assume then $n_1 \geq m_2$. Thus the sign of D is the same as that of

$$\begin{vmatrix} \frac{(n_1-m_1+b)!}{(n_1-m_1)!} & \frac{(n_1-m_2+b)!}{(n_1-m_2)!} \\ \frac{(n_2-m_1+b)!}{(n_2-m_1)!} & \frac{(n_2-m_2+b)!}{(n_2-m_2)!} \end{vmatrix},$$

and so is the same as that of

$$\left(1+\frac{m_2-m_1}{n_1-m_2+b}\right)\left(1+\frac{m_2-m_1}{n_1-m_2+b-1}\right)\cdots\left(1+\frac{m_2-m_1}{n_1-m_2+1}\right) - \left(1+\frac{m_2-m_1}{n_2-m_2+b}\right)\left(1+\frac{m_2-m_1}{n_2-m_2+b-1}\right)$$
$$\cdots\left(1+\frac{m_2-m_1}{n_2-m_2+1}\right),$$

which is positive.

QED

In a similar fashion we may treat the general case in which the failure densities are known with uncertainty. However, before using Procedure 1 or Procedure 2 for computing optimal kits, we must verify that the resulting $Q_i(n)$, the probability of at most n failures of the i^{th} type, has the property that $Q_i(n+1)/Q_i(n)$ is a decreasing function of n.

3. MULTIPERIOD INVENTORY POLICY

In Section 2, a mathematical solution is obtained to the problem of determining the composition of the spare-parts kit that achieves, at a minimum cost for spares, a specified protection against system shutdown due to shortage during a specified period. In Section 2.5.1, it is assumed further that the life of each component follows an exponential distribution with a known mean.

In the present section, we consider a model in which periodic resupply of the spares kit occurs. Here we wish to determine that resupply policy which achieves, at a minimum total cost, a specified protection during each period against system shutdown due to shortage [22]. The exact model is as follows:

3.1 Mathematical Model

A complex system is scheduled to operate in the field for a total length of time T_0. We wish to determine the inventory policy governing periodic resupply of spares for essential components which will yield assurance $P^{(p)} \geq a$ of continued system operation during period p, of length T, for p = 1, 2, ..., N, at minimum total discounted cost

$$L_N = \sum_{p=1}^{N} \alpha^{p-1} c^{(p)}, \tag{1}$$

where $0 < \alpha \leq 1$ is the discount factor (discussed in comment A below) and $c^{(p)}$ is the expenditure for spares during the p^{th} period, p = 1, 2, ..., N.

We assume:

(a) The length of life of component i, j (or its replacement) is an independent random variable with density $\mu_{ij} e^{-\mu_{ij} t}$ (μ_{ij}, the failure rate, is the reciprocal of the mean life). j = 1, 2, ..., d_i indexes components of the same type, while i = 1, 2, ..., k indexes the different component types.

(b) During period p, p = 1, 2, ..., N, component i, j is scheduled to operate t_{ij} hours, j = 1, 2, ..., d_i, i = 1, 2, ..., k.

(c) Replacement of a failed component is made immediately if a spare is available at the field site. If a spare is not available at the field site, an emergency spare is borrowed from another site and repaid as soon as possible. Such emergencies are considered very undesirable, since the system is shut down for a short time.

(d) At the beginning of the p^{th} period, p = 1, 2, ..., N, an inventory is taken of the number, $x_i^{(p)}$, of spares on hand of the i^{th} component type, i = 1, 2, ..., k. An order is then placed for $u_i^{(p)}$ spares of type i, i = 1, 2, ..., k, and delivery is made immediately. (In Section 3.3 we consider the more realistic case of lag in delivery.)

(e) The total cost, $c^{(p)}$, expended on spares during the p^{th} period, p = 1, 2, ..., N, is given by

$$c^{(p)} = \sum_{i=1}^{k} u_i^{(p)} c_i. \tag{2}$$

We seek an inventory policy, i.e., a set of functions $u_i^{(p)}\left(x_1^{(p)}, \ldots, x_k^{(p)}\right) \geq 0$, i = 1, 2, ..., k; p = 1, 2, ..., N; which minimize the expected cost EL_N subject to

$$P^{(p)} \geq a, \quad p = 1, 2, \ldots, N. \tag{3}$$

A few comments on this model are in order.

(A) The use of the discount factor α reflects directly the effect of an interest rate and indirectly the various holding costs, such as those due to storage, obsolescence, damage, etc. However, we shall see later that the optimal policy turns out to be independent of the value of α.

(B) No penalty cost is explicitly incorporated into the model. Rather, the occurrence of system shutdowns is controlled by demanding assurance $P^{(p)} \geq a$, $p = 1, 2, \ldots, N$ (i.e., the risk of system shutdown due to shortage during a specified period is required to be at most 1 - a). This seems to be a reasonable procedure in the military supply situation since it is intrinsically difficult to assign a penalty cost for the danger resulting from system shutdown. Thus, what we attempt to achieve in our model is an inventory policy which will buy adequate protection at minimum total expected cost.

(C) Note that the actual amounts to be ordered at the beginning of the p^{th} period are not decided at the beginning of the first period. Instead, the decision as to the actual amounts to be ordered is made at the beginning of the p^{th} period based on the stocks actually on hand at that time. What is determined in advance is the policy to be followed (i.e., how to use the stock-level data in determining the size of the order). Explicitly, the order size, $u_i^{(p)}$, for the i^{th} component type is a function of $x_1^{(p)}, \ldots, x_k^{(p)}$.

(D) The failure rates, μ_{ij}, are assumed to be known at the beginning. If, as is likely to be true in actual practise, the failure rate is estimated and the estimate becomes more reliable as additional information is accumulated, it would be wise to revise the values used for the μ_{ij} from time to time.

(E) If a spare must be borrowed from another site to replace a failed component, we shall treat the short time required to obtain the replacement as negligible in deriving the solution. However, the danger resulting from system shutdown during this short time is not negligible; thus, the requirement that the risk of experiencing such a shutdown during the p^{th} period be $\leq 1-a$, for $p = 1, 2, \ldots, N$.

(F) The present multiperiod model is somewhat more sensitive to the assumption of exponential failure densities than the one-period model treated previously. Since the total length of time involved will tend to be greater than in the one-period model, wearout failures, which are not characteristic of the exponential density (as distinquished from catastrophic failures, which are characteristic of the exponential density), will have a greater chance to occur.

3.2 Derivation of Optimal Policy

In this section we derive the optimal ordering policy (i.e., the form of the functions $u_i^{(p)}\left(x_1^{(p)}, \ldots, x_k^{(p)}\right)$, $i = 1, 2, \ldots, k$; $p = 1, 2, \ldots, N$; which minimize EL_N subject to (3) of Section 3.1).

We shall find it convenient to use vectors:

$$\underline{c} = (c_1, c_2, \ldots, c_k),$$

$$\underline{x}^{(p)} = (x_1^{(p)}, x_2^{(p)}, \ldots, x_k^{(p)}),$$

$$\underline{u}^{(p)} = (u_1^{(p)}, u_2^{(p)}, \ldots, u_k^{(p)}),$$

and, in general, an underlined quantity will represent a vector. Next, let $\underline{n}^*(T)$ represent the spare-parts kit yielding assurance $\geq a$ of continued system operation during a single period of length T at minimum cost. That is, $\underline{n}^*(T)$ represents the optimal kit of the single-period model treated in Section 2.5. Then we may express the optimal ordering policy for the present multiperiod model very simply in terms of $\underline{n}^*(T)$:

THEOREM 1: The optimal policy in the N-period inventory model is to order up to $\underline{n}^*(T)$ at the beginning of the p^{th} period, $p = 1, 2, \ldots, N$ (i.e., $\underline{u}^{(p)*}\left[\underline{x}^{(p)}\right] = \underline{n}^*(T) - \underline{x}^{(p)}$ minimizes EL_N subject to (3) of Section 3.1).

PROOF: Let $\underline{z}^{(p)} = (z_1^{(p)}, z_2^{(p)}, \ldots, z_k^{(p)})$ = the vector of numbers of spares of the k component types consumed during the p^{th} period, with $\underline{z}^{(0)} = \underline{0}$. Then, by ordering in accordance with the proposed policy, we expend

$$L_N^* = \underline{c} \cdot \left\{ \underline{n}^* + \sum_{j=1}^{N-1} \alpha^j \underline{z}^{(j)} \right\}, \tag{1}$$

where the dot signifies inner product. Note that (1) simply represents the cost expended in initially provisioning $\underline{n}^*$ (for convenience we drop the argument T) plus the discounted cost expended during each successive period in replacing the spares consumed during the previous period.

By ordering in accordance with any other decision function, $\underline{u}^{(p)}$, we expend

$$L_N = \underline{c} \cdot \left\{ \underline{u}^{(1)} + \alpha \underline{u}^{(2)} + \ldots + \alpha^{N-1} \underline{u}^{(N)} \right\}.$$

Define, for $p = 1, 2, \ldots, N$,

$$\underline{m}^{(p)} = \underline{x}^{(p)} + \underline{u}^{(p)};$$

thus $\underline{m}^{(p)}$ represents the vector of stock levels after delivery at the beginning of the p^{th} period. Since $\underline{u}^{(1)} = \underline{m}^{(1)}$ and $\underline{u}^{(p)} = \underline{m}^{(p)} - \underline{m}^{(p-1)} + \underline{z}^{(p-1)}$ for $p = 2, 3, \ldots, N$, we may write

$$L_N = \underline{c} \cdot \left\{ \underline{m}^{(1)} + \alpha(\underline{m}^{(2)} - \underline{m}^{(1)} + \underline{z}^{(1)}) + \alpha^2(\underline{m}^{(3)} - \underline{m}^{(2)} + \underline{z}^{(2)}) \right.$$
$$\left. + \ldots + \alpha^{N-1}(\underline{m}^{(N)} - \underline{m}^{(N-1)} + \underline{z}^{(N-1)}) \right\}.$$

Regrouping terms, we get

$$L_N = \underline{c} \cdot \left\{ (1-\alpha) \left[\underline{m}^{(1)} + \alpha \underline{m}^{(2)} + \alpha^2 \underline{m}^{(3)} + \ldots + \alpha^{N-2} \underline{m}^{(N-1)} \right] \right.$$
$$\left. + \alpha^{N-1} \underline{m}^{(N)} + \sum_{j=1}^{N-1} \alpha^j \underline{z}^{(j)} \right\}$$

$$\geq \underline{c} \cdot \left\{ (1-\alpha) \left(\underline{n}^* + \alpha \underline{n}^* + \ldots + \alpha^{N-2} \underline{n}^* \right) + \alpha^{N-1} \underline{n}^* + \sum_{j=1}^{N-1} \alpha^j \underline{z}^{(j)} \right\}$$

since $\underline{n}^*$ is the minimum cost vector of stock levels yielding assurance $\geq a$ of continued system operation during a single period. Hence, from (1),

$$L_N \geq \underline{c} \cdot \left\{ (1-\alpha)\frac{1-\alpha^{N-1}}{1-\alpha}\underline{n}^* + \alpha^{N-1}\underline{n}^* + \sum_{j=1}^{N-1} \alpha^j \underline{z}(j) \right\}$$

$$= \underline{c} \cdot \left\{ \underline{n}^* + \sum_{j=1}^{N-1} \alpha^j \underline{z}(j) \right\} = L_N^*$$

Hence

$$EL_N \geq EL_N^*.$$

QED

Theorem 1 tells us that the optimal policy for the multiperiod model is to order initially the optimal vector of spares for the one-period problem (Sec. 2.5), and then during each subsequent period simply to order as many spares of the i^{th} type, $i = 1, 2, \ldots, k$, as there were failures during the previous period. Thus, the optimal policy is intuitively appealing and administratively very easy to follow.

The expected cost, EL_N^*, of operating the optimal policy for N periods is given by

$$EL_N^* = \underline{c} \cdot \underline{n}^* + \underline{c} \cdot E \sum_{j=1}^{N-1} \alpha^j \underline{z}(j) = \underline{c} \cdot \underline{n}^* + \frac{1-\alpha^N}{1-\alpha}\, \underline{c} \cdot E\, \underline{z},$$

where $\underline{z} = (z_1, z_2, \ldots, z_k)$, with the z_i independently distributed according to the Poisson distribution with parameter

$$\lambda_i = \sum_{j=1}^{d_i} \mu_{ij}\, t_{ij}.$$

Thus

$$EL_N^* = \underline{c} \cdot \underline{n}^* + \frac{1-\alpha^N}{1-\alpha}\, \underline{c} \cdot \lambda, \tag{2}$$

where $\underline{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_k)$. Note that as $N \to \infty$,

$$EL_N^* \to \underline{c} \cdot \underline{n}^* + \frac{1}{1-\alpha}\underline{c} \cdot \underline{\lambda};$$

this represents the total discounted cost of operating the optimal policy over an infinite length of time.

3.3 Lag in Delivery

Consider the same model as in Section 3.1, but assume now that delivery takes one period. Let $\underline{u}^{(p)}$ now represent the order to be delivered at the beginning of the p^{th} period (the order having been placed at the beginning of the p-1^{st} period), p = 1, 2, . . . , N. Then we may state:

THEOREM 1: The optimal policy for the N period inventory model with one-period lag in delivery is given by $\underline{u}^{(1)} = \underline{n}^*(T)$, $\underline{u}^{(p+1)} = \underline{n}^*(2T) - \underline{m}^{(p)}$, p = 1, 2, . . . , N-1.

PROOF: The proof consists of three parts. First, we verify that $\underline{u}^{(2)}$ as proposed is non-negative; although intuitively very reasonable, this statement requires a number of mathematical steps. Second, we verify that we obtain the desired protection under the proposed policy. Finally, we verify that its expected cost is minimum among competing policies.

We must first verify that $\underline{u}^{(2)} = \underline{n}^*(2T) - \underline{n}^*(T) \geq 0$. Let

$\underline{\lambda}(T)$ = vector of expected number of failures during a period of length T.

Then

$$\underline{\lambda}(2T) = \underline{2}\lambda(T).$$

Let

$P_i^{(T)}(n)$ = probability of experiencing $\leq$ n failures of the i^{th} component type during a period of length T.

Then

$$P_i^{(T)}(n) = \sum_{x=0}^{n} e^{-\lambda_i(T)} \frac{[\lambda_i(T)]^x}{x!}$$

since the failure densities are exponential. (See Section 2.5.1.) Hence, omitting the argument of $\lambda_i(T)$, we get for i = 1, 2, . . . , k,

$$\frac{P_i^{(T)}(n+1)}{P_i^{(T)}(n)} = 1 + \frac{\frac{\lambda_i^{n+1}}{(n+1)!}}{\sum_{x=0}^{n} \lambda_i^x/x!} \leq 1 + \frac{\frac{(2\lambda_i)^{n+1}}{(n+1)!}}{\sum_{x=0}^{n} (2\lambda_i)^x/x!}$$

$$= \frac{P_i^{(2T)}(n+1)}{P_i^{(2T)}(n)}.$$

From this it follows that

$$\ln P_i^{(T)}(n+1) - \ln P_i^{(T)}(n) \leq \ln P_i^{(2T)}(n+1) - \ln P_i^{(2T)}(n), \quad i = 1, 2, \ldots, k.$$

Thus,

$$\text{either } n_i^*(T) \geq n_i^*(2T) \quad \text{for} \quad i = 1, 2, \ldots, k,$$

$$\text{or} \quad n_i^*(T) \leq n_i^*(2T) \quad \text{for} \quad i = 1, 2, \ldots, k.$$

But $n_i^*(T) \geq n_i^*(2T)$ for all i with strict inequality for at least one i is impossible, since if $\underline{n}^*(2T)$ yields assurance $\geq a$ for a period of length 2T, $\underline{n}^*(2T)$ will yield assurance $\geq a$ for a period of length T. Thus $\underline{n}^*(2T)$ would provide the required assurance at lesser cost than $\underline{n}^*(T)$, contradicting the fact that $\underline{n}^*(T)$ is, by definition, the least cost vector of stock levels providing assurance $\geq a$ against shortage during the period of length T. Hence we conclude that $\underline{n}^*(2T) \geq \underline{n}^*(T)$, so that $\underline{u}^{(2)} \geq 0$.

Next we must verify that during each period we obtain the specified protection $\geq a$ against system shutdown due to shortage. For the first period, we start with a kit of $\underline{n}^*(T)$ which by construction yields a protection $\geq a$. A moment's reflection shows that under any policy yielding probability $\geq a$ of no shortage during the p^{th} period, p = 1, 2, ..., N, we have

$$(1) \qquad P\left[\underline{z}^{(p-1)} + \underline{z}^{(p)} \leq \underline{M}^{(p-1)}\right] \geq a,$$

where

$$\underline{M}^{(0)} = \underline{m}^{(1)}, \text{ and } \underline{M}^{(p)} = \underline{m}^{(p)} + \underline{u}^{(p+1)}, \quad p = 1, 2, \ldots, N-1;$$

that is, no shortage during the p^{th} period will occur if, and only if, the total demand during the $p\text{-}1^{st}$ and p^{th} periods does not exceed, for any component type, the total of spares available during these two periods. Thus, the probability of no shortage during the p^{th} period under the proposed policy is $P\left[\underline{z}^{(p-1)} + \underline{a}^{(p)} \leq \underline{n}^*(2T)\right] \geq a$, since $\underline{z}^{(p-1)} + \underline{z}^{(p)}$ represents the total demand during an interval of length 2T, and $\underline{M}^{(p-1)} = \underline{m}^{(p-1)} + \underline{u}^{(p)} = \underline{n}^*(2T)$, p = 2, 3, ..., N.

Finally, we must show that the expected cost EL_N^* of the proposed policy is minimum among all inventory policies yielding the required protection. First note that under any policy

$$(2) \qquad EL_N = E\, \underline{c} \cdot \left\{ \underline{u}^{(1)} + \alpha \underline{u}^{(2)} + \alpha^2 \underline{u}^{(3)} + \ldots + \alpha^{N-1} \underline{u}^{(N)} \right\}.$$

Next note that

$$(3) \qquad \underline{u}^{(p+1)} = \underline{M}^{(p)} - \underline{M}^{(p-1)} + \underline{z}^{(p-1)}, \quad p = 1, 2, \ldots, N-1.$$

From (2) and (3), we get

$$EL_N = E\, \underline{c} \cdot \left\{ \underline{m}^{(1)} + \sum_{p=1}^{N-1} \alpha^p \left[\underline{M}^{(p)} - \underline{M}^{(p-1)} + \underline{z}^{(p-1)} \right] \right\},$$

or

$$(4) \qquad EL_N = \underline{c} \cdot \left\{ (1-\alpha)\left[\underline{m}^{(1)} + \alpha \underline{M}^{(1)} + \ldots + \alpha^{N-2} \underline{M}^{(N-2)}\right] + \alpha^{N-1} \underline{M}^{(N-1)} + \sum_{p=1}^{N-2} \alpha^{p+1} E\, \underline{z}^{(p)} \right\}.$$

Now the proposed policy is equivalent to specifying $\underline{m}^{(1)} = \underline{n}^*(T)$, $\underline{M}^{(p)} = \underline{n}^*(2T)$ for $p = 1, 2, \ldots, N-1$. Hence it minimizes EL_N, since $\underline{n}^*(T)$ minimizes $\underline{c} \cdot \underline{m}^{(1)}$ for $\underline{m}^{(1)}$ satisfying $P[\underline{z}^{(1)} \leq \underline{m}^{(1)}] \geq a$, while $\underline{n}^*(2T)$ minimizes $\underline{c}\, \underline{M}^{(p)}$ for $\underline{M}^{(p)}$ satisfying $P[\underline{z}^{(p)} + \underline{z}^{(p+1)} \leq \underline{M}^{(p)}] \geq a$, $p = 1, 2, \ldots, N-1$; the final sum of (4) is, of course, independent of policy.

QED

It is clear how to generalize Theorem 1 to the case of lag in delivery of L periods, $L < N$, where we assume, however, that it is possible at the beginning of periods 2, 3, ..., L to arrange for delivery of the amounts specified at the very beginning. Then we may establish by arguments similar to those of Theorem 1:

THEOREM 2: The optimal policy in the N period inventory model where delivery lag is L periods, is $\underline{u}^{(1)} = \underline{n}^*(T)$, $\underline{u}^{(2)} = \underline{n}^*(2T) - \underline{n}^*(T)$, ..., $\underline{u}^{(L+1)} = \underline{n}^*[(L+1)T] - \underline{n}^*(LT)$; $\underline{u}^{(L+2)} = \underline{z}^{(1)}$, $\underline{u}^{(L+3)} = \underline{z}^{(2)}$, ..., $\underline{u}^{(N)} = \underline{z}^{(N-L-1)}$.

Our final extension is to the case where the lag in delivery is a fraction f, $0 < f < 1$, of the period-length T. We assume an order is placed at time pT and delivered at time $pT + f$, $p = 0, 1, 2, \ldots, N-1$. For convenience, we require separate assurance $\geq a$ of continued system operation during each of the periods $[0, fT)$, $[fT, (1+f)T)$, ..., $[(N-1+f)T, (N+f)T]$.

Again, we may use arguments similar to those of Theorem 2 to establish:

THEOREM 3: The optimal policy when the delivery lag is fT, $0 < f < 1$, is $\underline{u}^{(1)} = \underline{n}^*(fT)$, $\underline{u}^{(1+f)} = \underline{n}^*[(1+f)T] - \underline{n}^*(fT)$; $\underline{u}^{(p+f)} = \underline{z}^{(p-1)}$ for $p = 2, 3, \ldots, N$, where $\underline{u}^{(1)}$ = vector of stock provided at time 0, $\underline{u}^{(p+f)}$ = vector of stock scheduled for delivery at time $(p-1)T + f$, and $\underline{z}^{(p)}$ = vector of numbers of failures experienced during the interval $(p-1)T$ to pT, $p = 1, 2, \ldots, N$.

In a similar fashion, one can solve the case of lag of any length.

3.4 Budgeting for Spares Expenditures

In some applications, it is necessary to budget sufficient funds to cover the cost of spares for, say, b periods. Since the actual number of failures of each component type is a random variable, we cannot set aside funds which will exactly cover spares expenditures; instead, we wish to provide an amount B which will have a specified high degree of assurance, β, of covering spares expenditures.

First note that the actual total expenditure D, ignoring discounting, is given by

$$D = \sum_{i=1}^{k} \sum_{p=1}^{b} c_i z_i^{(p)} = \sum_{i=1}^{k} c_i y_i$$

where y_i, the total number of spares of the i^{th} type consumed during the b periods, is a Poisson random variate with mean $b\lambda_i$. Thus

$$ED = b \sum_{i=1}^{k} c_i \lambda_i,$$

and

$$\sigma_D^2 = b \sum_{i=1}^{k} c_i^2 \lambda_i,$$

since the variance of a Poisson random variate is the same as its mean.

Now the exact distribution of D will in general be complicated. However, since in practical applications k will be large, by the Central Limit Theorem we may approximate the distribution of D by the normal distribution. Thus we conclude that the amount to budget is approxmately given by

$$B = b \sum_{i=1}^{k} c_i \lambda_i + t_\beta \sqrt{b \sum_{i=1}^{k} c_i^2 \lambda_i} \quad , \tag{1}$$

where t_β is the β^{th} percentile of the normal distribution with mean 0, variance 1.

APPENDIX

4. POLYA-TYPE DISTRIBUTIONS of CONVOLUTIONS

4.1 Introduction

The theory of totally positive kernels and Polya-type distributions has been extensively applied in several domains of mathematics, statistics, economics, and mechanics. Totally positive kernels arise naturally in developing procedures for inverting, by differential polynomial operators [9], integral transformations defined in terms of convolution kernels. The theory of Polya-type distributions are fundamental in permitting characterizations of best statistical procedures for decision problems [14, 15, 18]. In clarifying the structure of stochastic processes with continuous path functions, we encounter totally positive kernels [16]. Studies in the stability of certain models in mathematical economics frequently use properties of totally positive kernels [11]. The theory of vibrations of certain types of mechanical systems involves aspects of the theory of totally positive kernels [5].

In this appendix (See also [17]) we characterize new classes of totally positive kernels that arise from summing independent random variables and forming related first passage time distributions. The results are used in obtaining the sufficient condition for Procedures 1 and 2 of Section 2.2.1.

A function f(x, y) of two real variables ranging over linearly ordered one-dimensional sets X and Y, respectively, is said to be totally positive of order k (TP_k) if for all $x_1 < x_2 < \ldots < x_m$, $y_1 < y_2 < \ldots < y_m$, (x_i in X; y_j in Y), and all $1 \leq m \leq k$,

$$(1)\qquad f\begin{pmatrix} x_1, & x_2, & \cdots, & x_m \\ y_1, & y_2, & \cdots, & y_m \end{pmatrix} = \begin{vmatrix} f(x_1,y_1) & f(x_1,y_2) & \cdots & f(x_1,y_m) \\ f(x_2,y_1) & f(x_2,y_2) & \cdots & f(x_2,y_m) \\ \vdots & \vdots & \ddots & \vdots \\ f(x_m,y_1) & f(x_m,y_2) & \cdots & f(x_m,y_m) \end{vmatrix} \geq 0.$$

Typically, X is an interval of the real line, or a countable set of discrete values on the real line, such as the set of all integers or the set of nonnegative integers; similarly for Y. When X or Y is a set of integers, we may use the term "sequence" rather than "function."

A related, weaker property is that of sign regularity. A function $f(x, y)$ is sign regular of order k, if for every $x_1 < x_2 < \cdots < x_m$, $y_1 < y_2 < \cdots < y_m$, and $1 \leq m \leq k$, the sign of

$$f\begin{pmatrix} x_1, & x_2, & \cdots, & x_m \\ y_1, & y_2, & \cdots, & y_m \end{pmatrix}$$

depends on m alone.

If a TP_k function $f(x, y)$ is a probability density in one of the variables, say x, with respect to a σ-finite measure $\mu(x)$, for each fixed value of y, then $f(x, y)$ is said to be Polya type of order k (PT_k). The concepts of PT_1 and PT_2 densities are familiar ones. Every density characterized by a parameter is PT_1; while the PT_2 densities are those having a monotone likelihood ratio [18].

A further specialization occurs if a PT_k kernel may be written as a function $f(x-y)$ of the difference of x and y, where x and y traverse the real line; $f(u)$ is then said to be a Polya frequency density of order k (PF_k).

Finally, if the subscript ∞ is written in any of the definitions, then the property in question will be understood to hold for all positive integers.

Many of the structural properties of TP_k functions are deducible from the following identity, which appears in [21, p. 48, problem 68]:

LEMMA 1: If $r(x, w) = \int p(x, t)\, q(t, w)\, d\sigma(t)$, then

$$(2)\qquad r\begin{pmatrix} x_1, & x_2, & \cdots, & x_k \\ w_1, & w_2, & \cdots, & w_k \end{pmatrix} = \int_{t_1 < t_2 < \cdots < t_k} \cdots \int p\begin{pmatrix} x_1, & x_2, & \cdots, & x_k \\ t_1, & t_2, & \cdots, & t_k \end{pmatrix} q\begin{pmatrix} t_1, & t_2, & \cdots, & t_k \\ w_1, & w_2, & \cdots, & w_k \end{pmatrix} d\sigma(t_1)\, d\sigma(t_2) \cdots d\sigma(t_k).$$

In particular, we obtain from Lemma 1:

LEMMA 2: If $f(x, t)$ is TP_m and $g(t, w)$ is TP_n, then $h(x, w) = \int f(x, t)\, g(t, w)\, d\sigma(t)$ (the convolution of f and g) is $TP_{\min(m, n)}$.

We shall exploit this result principally in the case when f and g are Polya frequency densities: Therefore,

LEMMA 3: If $f(x-t)$ is PF_m and $g(t-w)$ is PF_n, then $h(x-w) = \int f(x-t)\, g(t-w)\, dt$ is $PF_{\min(m, n)}$.

An important feature of totally positive functions is their variation-diminishing property: If $f(x,w)$ is TP_k and $g(w)$ changes sign at most $j \leq k-1$ times, then $h(x) = \int f(x,w)\, g(w)\, d\sigma(w)$ changes sign at most j times; moreover, if $h(x)$ actually changes sign j times, then it must change sign in the same order as $g(w)$ [14, 15]. This distinctive property underlies many of the applications mentioned above. The variation-diminishing property is essentially equivalent to the determinantal inequalities (1).

4.2 Convolution of Nonnegative Random Variables

We first prove

THEOREM 1: Let $f_1, f_2, \ldots$ be of any sequence of densities of nonnegative random variables, with each f_i a PF_k. Then the n-fold convolution $g(n,x) = f_1 * f_2 * \ldots * f_n(x)$ is PT_k in the variables n and x, where n ranges over 1, 2, ... and x traverses the positive real line.

PROOF: The proof employs induction. First note that $g(n,x)$ is PT_1 since $g(n,x) \geq 0$ for each real x and each positive integer n.

Assume now that for every sequence of densities satisfying the hypothesis, the corresponding n-fold convolution has been proven PT_{r-1} for $r \leq k$. We prove that this implies that $g(n,x)$ is PT_r.

(a) First consider the case $n_1 = 1$. Given $1 < n_2 < n_3 < \ldots < n_r$, $0 \leq x_1 < x_2 < \ldots < x_r$, we may write

$$(1) \qquad g\begin{pmatrix} 1, & n_2, & \ldots, & n_r \\ x_1, & x_2, & \ldots, & x_r \end{pmatrix} = \sum_{\nu=1}^{r} (-1)^{\nu-1} f_1(x_\nu)\, g\begin{pmatrix} n_2 & n_3, & \ldots & , n_r \\ x_1, \ldots, x_{\nu-1}, & x_{\nu+1}, & \ldots, & x_r \end{pmatrix}$$

simply by expanding the determinant on the left by its first row. Next note that for $n = 2, 3, \ldots$ and $x \geq 0$,

$$(2) \qquad g(n,x) = \int g_1(n-1,\xi)\, f_1(x-\xi)\, d\xi,$$

where $g_1(n-1,\xi)$ is defined as $f_2 * f_3 * \ldots * f_n(\xi)$. Inserting (2) of Section 4.1 in (2) just above, we may write

$$(3) \qquad g\begin{pmatrix} n_2, n_3, \ldots & , n_r \\ x_1, \ldots, x_{\nu-1}, x_{\nu+1}, \ldots, & x_r \end{pmatrix} = \int_{0 \leq \xi_1 < \xi_2 < \ldots < \xi_{r-1}} \cdots \int g_1\begin{pmatrix} n_2-1, n_3-1, \ldots, & n_r-1 \\ \xi_1, \xi_2, \ldots & , \xi_{r-1} \end{pmatrix} f_1\begin{pmatrix} x_1, \ldots, x_{\nu-1}, x_{\nu+1}, \ldots, & x_r \\ \xi_1, \xi_2, \ldots & , \xi_{r-1} \end{pmatrix} d\xi_1, d\xi_2, \ldots, d\xi_{r-1}.$$

Inserting (3) into (1), we get immediately

$$g\begin{pmatrix} 1, n_2, \dots, n_r \\ x_1, x_2, \dots, x_r \end{pmatrix} = \int_{0 < \xi_1 < \xi_2 < \dots < \xi_{r-1}} \cdots \int g_1\begin{pmatrix} n_2-1, n_3=1, \dots, n_r-1 \\ \xi_1, \xi_2, \dots \quad , \xi_{r-1} \end{pmatrix}$$

$$\sum_{\nu=1}^{r} (-1)^{\nu-1} f_1(x_\nu) f_1\begin{pmatrix} x_1, \dots, x_{\nu-1}, x_{\nu+1}, \dots, x_r \\ \xi_1, \xi_2, \dots \quad , \xi_{r-1} \end{pmatrix} d\xi_1, d\xi_2, \dots, d\xi_{r-1}$$

(4)

$$= \int_{0 \le \xi_1 < \xi_2 < \dots < \xi_{r-1}} \cdots \int g_1\begin{pmatrix} n_2-1, n_3-1, \dots, n_r-1 \\ \xi_1, \xi_2, \dots \quad , \xi_{r-1} \end{pmatrix}$$

$$\cdot f_1\begin{pmatrix} x_1, x_2, \dots \quad , x_r \\ 0, \xi_1, \xi_2, \dots, \xi_{r-1} \end{pmatrix} d\xi_1, d\xi_2, \dots, d\xi_{r-1}.$$

But

$$g_1\begin{pmatrix} n_2-1, & n_3-1, & \dots, & n_r-1 \\ \xi_1 & , \xi_2 & , \dots, & \xi_{r-1} \end{pmatrix} \ge 0$$

by the induction assumption, while

$$f_1\begin{pmatrix} x_1, & x_2, & \dots, & x_r \\ 0, & \xi_1, & \xi_2, & \dots, & \xi_{r-1} \end{pmatrix} \ge 0$$

since f_1 is Pff_k by the hypothesis of the theorem since $0 \le \xi_1 < \xi_2 < \dots < \xi_{r-1}$. Hence

(5)
$$g\begin{pmatrix} 1, & n_2, & \dots, & n_r \\ x_1, & x_2, & \dots, & x_r \end{pmatrix} \ge 0.$$

(b) Now suppose $n_1 > 1$. Then for any $n_1 < n_2 < \ldots < n_k$ and $0 \leq x_1 < x_2 < \ldots < x_k$, we may write, using (2) of Section 4.1 and (2) of the present section,

$$g\begin{pmatrix} n_1, & n_2, & \ldots, & n_r \\ x_1, & x_2, & \ldots, & x_r \end{pmatrix} = \int \cdots \int_{\xi_1 < \xi_2 < \ldots < \xi_r} g_1\begin{pmatrix} n_1-1, & n_2-1, & \ldots, & n_r-1 \\ \xi_1 & , \xi_2 & , \ldots, & \xi_r \end{pmatrix}$$

(6)

$$\cdot\, f_1\begin{pmatrix} x_1, & x_2, & \ldots, & x_r \\ \xi_1, & \xi_2, & \ldots, & \xi_r \end{pmatrix} d\xi_1, d\xi_2, \ldots, d\xi_r .$$

From (6) we see that for every sequence of densities satisfying the hypothesis, the corresponding functions g_1, g satisfy

$$g_1\begin{pmatrix} n_1-1, & n_2-1, & \ldots, & n_r-1 \\ \xi_1 & , \xi_2 & , \ldots, & \xi_r \end{pmatrix} \geq 0$$

implies

$$g\begin{pmatrix} n_1, & n_2, & \ldots, & n_r \\ x_1, & x_2, & \ldots, & x_r \end{pmatrix} \geq 0. \tag{7}$$

By use of (5) and (7), it follows by induction that

$$g\begin{pmatrix} n_1, & n_2, & \ldots, & n_r \\ x_1, & x_2, & \ldots, & x_r \end{pmatrix} \geq 0.$$

Since $g(n, x)$ has thereby been proved to be PT_r, we have established the validity of the induction step, and the theorem follows.

QED

It is important to emphasize the distinction between Lemma 3 of Section 4.1 and Theorem 1 of this section. Under the hypothesis of Theorem 1, Lemma 3 of Section 4.1 states that for each fixed positive integer n, $g(n, x)$ is PT_k in differences of x, while Theorem 1 states that $g(n, x)$ is PT_k in the variables n and x.

4.3 Random Variables Over the Entire Real Line

Will Theorem 1 of Section 4.2 hold if the random variables are not restricted to be nonnegative? In general, the answer is no, as the following example shows.

EXAMPLE: Let $f_1(x) = f_2(x) = \ldots = 1\, e^{-x^2/2}/\sqrt{2\pi}$, a PF_∞. Then $g(n,x) = f^{(n)}(x) = 1\, e^{-x^2/2n}/\sqrt{2\pi}$. For $1 \leq n_1 \leq n_2$, $x_1 < x_2$, the second order determinant is positive for $0 \leq x_1 < x_2$ and negative for $x_1 < x_2 \leq 0$. Thus $g(n,x)$ is not PT_2.

However, a generalization of Theorem 1 of Section 4.2 to the case of random variables ranging over the whole real line is possible, as developed in Theorem 1 below. In the more general case, total positivity applies, not to the n-fold convolution, but rather to the first passage time probabilities of the partial sums.

THEOREM 1: Let $f_1, f_2, \ldots$ be any sequence of PT_k densities of random variables $X_1, X_2, \ldots$, respectively, which are not necessarily nonnegative. Consider the first passage probability for x positive:

$$h(n,x) = P\left(\sum_{i=1}^{n} X_i \geq x;\ \sum_{i=1}^{j} X_i < x,\ j = 1, 2, \ldots, n-1\right) \quad \text{for } n = 1, 2, \ldots .$$

Then $h(n,x)$ is TP_k, where n ranges over 1, 2, ... and x traverses the positive axis.

PROOF: The proof proceeds in a similar fashion to that of Theorem 1 of Section 4.2. We employ induction. First we note that $h(n,x)$ is TP_1 since $h(n,x) \geq 0$ for $n = 1, 2, \ldots$ and each positive x.

Assume now that for every sequence of densities satisfying the hypothesis the associated first passage time probability function is TP_{r-1} for $r \leq k$. We shall prove that this implies $h(n,x)$ is TP_r. From this the conclusion of the theorem will follow.

We clearly have for x positive that

$$h(n,x) = \begin{cases} \displaystyle\int_x^\infty f(\xi)\,d\xi & \text{for } n = 1 \\ \displaystyle\int_0^\infty f(x-\xi)\, h_1(n-1,\xi)\,d\xi & \text{for } n \geq 2 \end{cases} \tag{1}$$

where

$$h_1(n-1,\xi) = P\left(\sum_{i=2}^{n} X_i \geq \xi;\ \sum_{i=2}^{j} X_i < \xi,\ j = 2, 3, \ldots, n-1\right)$$

(a) We consider first the case $n_1 = 1$. Given $1 < n_2 < n_3 < \ldots < n_r$, $x_1 < x_2 < \ldots < x_r$, we may write, using (1),

$$h\begin{pmatrix}1, n_2, n_3, \dots, n_r\\ x_1, x_2, x_3, \dots, x_r\end{pmatrix} = \begin{vmatrix} \int_0^\infty f_1(x_1+\xi)d\xi & \int_0^\infty f_1(x_2+\xi)d\xi & \dots & \int_0^\infty f_1(x_r+\xi)d\xi \\ h(n_2, x_1) & h(n_2, x_2) & \dots & h(n_2, x_r) \\ \vdots & \vdots & \ddots & \vdots \\ h(n_r, x_1) & h(n_r, x_2) & \dots & h(n_r, x_r) \end{vmatrix}$$

$$= \sum_{\nu=1}^{r} (-1)^{\nu-1} \int_0^\infty f_1(x_\nu+\xi)\, h\begin{pmatrix}n_2, n_3, \dots, n_r\\ x_1, \dots, x_{\nu-1}, x_{\nu+1}, \dots, x_r\end{pmatrix} d\xi .$$

By using (1) of the present Section and (2) of Section 4.1, we now obtain

$$h\begin{pmatrix}n_2, n_3, \dots, n_r\\ x_1, \dots, x_{\nu-1}, x_{\nu+1}, \dots, x_r\end{pmatrix}$$

$$(2) \qquad = \int_{0 \le \xi_1 < \xi_2 < \dots < \xi_{r-1}} f_1\begin{pmatrix}x_1, \dots, x_{\nu-1}, x_{\nu+1}, \dots, x_r\\ \xi_1, \xi_2, \dots, \xi_{r-1}\end{pmatrix}$$

$$\cdot\, h_1\begin{pmatrix}n_2-1, n_3-1, \dots, n_r-1\\ \xi_1, \xi_2, \dots, \xi_{r-1}\end{pmatrix} d\xi_1, d\xi_2, \dots, d\xi_{r-1} .$$

Inserting (2) in the equation above it and replacing $-\xi$ by ξ gives

$$h\begin{pmatrix}1, n_2, n_3, \dots, n_r\\ x_1, x_2, x_3, \dots, x_r\end{pmatrix} = \int_{\xi < 0 \le \xi_1 < \xi_2 < \dots < \xi_{r-1}} h_1\begin{pmatrix}n_2-1, n_3-1, \dots, n_r-1\\ \xi_1, \xi_2, \dots, \xi_{r-1}\end{pmatrix} \sum_{\nu=1}^{r} (-1)^{\nu-1}$$

$$\cdot\, f_1(x-\xi)\, f_1\begin{pmatrix}x_1, \dots, x_{\nu-1}, x_{\nu+1}, \dots, x_r\\ \xi_1, \xi_2, \dots, \xi_{r-1}\end{pmatrix} d\xi\, d\xi_1\, d\xi_2 \dots d\xi_{r-1}$$

$$= \int_{\xi < 0 \le \xi_1 < \xi_2 < \dots < \xi_{r-1}} h_1\begin{pmatrix}n_2-1, n_3-1, \dots, n_r-1\\ \xi_1, \xi_2, \dots, \xi_{r-1}\end{pmatrix}$$

$$\cdot\, f_1\begin{pmatrix}x_1, x_2, \dots, x_r\\ \xi, \xi_1, \xi_2, \dots, \xi_{r-1}\end{pmatrix} d\xi\, d\xi_1\, d\xi_2 \dots d\xi_{r-1} .$$

But

$$h_1\begin{pmatrix} n_2-1, n_3-1, \ldots, n_r-1 \\ \xi_1, \quad \xi_2, \ldots, \xi_{r-1} \end{pmatrix} \geq 0$$

by the inductive assumption, while

$$f_1\begin{pmatrix} x_1, x_2, \cdots \quad, x_r \\ \xi, \xi_1, \xi_2, \ldots, \xi_{r-1} \end{pmatrix} \geq 0$$

since $\xi < \xi_1 < \xi_2 < \ldots < \xi_{r-1}$, $x_1 < x_2 < \ldots < x_r$, and $f_1(x)$ is PF_k by hypothesis. Hence

$$h\begin{pmatrix} 1, n_2, n_3, \ldots, n_r \\ x_1, x_2, x_3, \ldots, x_r \end{pmatrix} \geq 0.$$

The remainder of the proof parallels the corresponding portion of the proof of Theorem 1 of Section 4.2; simply replace g by h.

It is appropriate to compare Theorem 1 of Section 4.2 and Theorem 1 of this section. For this purpose, we sketch an argument which shows that Theorem 1 of Section 4.2 is actually a limiting case of Theorem 1 of this section. A careful examination of the preceding argument reveals that, in the case of nonnegative PF_k random variables, the probability of first passage at time n into any positive interval, not just the interval $[x, \infty]$, is TP_k. With the interval shrunk to a point, it readily follows that the first passage time probability converges to the density corresponding to the n-fold convolution. Since total positivity is preserved under this limiting operation, Theorem 1 of Section 4.2 follows.

4.4 Consequences of the Theorems

We now develop a series of consequences of Theorem 1 of Section 4.2 and Theorem 1 of Section 4.3. Let $F_i(x)$ be the cumulative distribution function corresponding to $f_i(x)$, $i = 1, 2, \ldots$. Then as a direct corollary of Theorem 1 of Section 4.3 we have:

THEOREM 1: Under the assumptions of Theorem 1 of Section 4.2 $h(n,x) = F_1 * F_2 * \ldots * F_{n-1}(x) - F_1 * F_2 * \ldots * F_n(x)$ is TP_k, where n ranges over 1, 2, ... and $x > 0$. In particular, if $f_i = f$, $i = 1, 2, \ldots$, then $h(n,x) = F^{(n-1)}(x) - F^{(n)}(x)$ is TP_k.

PROOF: Simply note that

$$h(n,x) = P\left(\sum_{i=1}^{n} X_i \geq x; \ \sum_{i=1}^{j} X_i < x, \ j = 1, 2, \ldots, n-1\right)$$ since the random variables are nonnegative.

QED

Actually we can say more about $h(n,x)$ in the situation where the X_i are nonnegative, independent, and identically distributed random variables; Theorem 2 just below asserts that for each fixed $x > 0$, $h(n+m,x)$ is sign regular in the variables $n \geq 0$ and $m \geq 0$.

THEOREM 2: Suppose $f(x)$ is PF_k with $f(x) = 0$ for $x < 0$. We define $h(n,x)$ by $h(n,x) = F^{(n-1)}(x) - F^{(n)}(x)$ for $n = 1, 2, \ldots$; $x \geq 0$, and for fixed $x \geq 0$ we define

$$c(n) = \begin{cases} h(n,x) & \text{for } n \geq 0 \\ 0 & \text{for } n \leq 0 . \end{cases}$$

Then $c(n+m)$ is sign regular of order k in $n \geq 1$ and $m \geq 1$; moreover, for $1 \leq n_1 < n_2 < \ldots < n_r$, $1 \leq m_1 < m_2 < \ldots < m_r$, the sign of

$$c_+\begin{pmatrix} n_1, & n_2, & \ldots, & n_r \\ m_1, & m_2, & \ldots, & m_r \end{pmatrix} \text{ is } (-1)^{r(r-1)/2},$$

where

$$c_+\begin{pmatrix} n_1, & n_2, & \ldots, & n_r \\ m_1, & m_2, & \ldots, & m_r \end{pmatrix} = \begin{vmatrix} c(n_1 + m_1) & \cdots & c(n_1 + m_r) \\ \vdots & & \vdots \\ c(n_r + m_1) & \cdots & c(n_r + m_r) \end{vmatrix} .$$

PROOF: For $m \geq 1$, we have

$$c(n+m) = \int g(m,\xi)\, h(n, x-\xi)\, d\xi \tag{1}$$

where $g(m,\xi) = f^{(m)}(\xi)$. Equation (1) simply states that if the partial sum first exceeds x at the $n+m^{th}$ stage, then this can occur by having the m^{th} partial sum equal to some nonnegative $\xi < x$, while the partial sum starting with the $r+1^{st}$ variable first exceeds $x-\xi$ at the n^{th} stage. From (1) of the present section and (2) of Section 4.1, we get, for $1 \leq n_1 < n_2 < \ldots < n_r$, $1 \leq m_1 < m_2 < \ldots < m_r$, $r \leq k$,

$$c_+\begin{pmatrix} n_1, & n_2, & \ldots, & n_r \\ m_1, & m_2, & \ldots, & m_r \end{pmatrix} = \int_0^{\infty} \int_{\xi_1}^{\infty} \cdots \int_{\xi_{r-1}}^{\infty} g\begin{pmatrix} m_1, & m_2, & \ldots, & m_r \\ \xi_1, & \xi_2, & \ldots, & \xi_r \end{pmatrix} \tag{2}$$

$$\cdot h\begin{pmatrix} n_1, & n_2, & \ldots, & n_r \\ x-\xi_1, & x-\xi_2, & \ldots, & x-\xi_r \end{pmatrix} d\xi_1\, d\xi_2 \cdots d\xi_r .$$

By Theorem 1 of Section 4.2,

$$g\begin{pmatrix} m_1, m_2, \cdots, m_r \\ \xi_1, \xi_2, \cdots, \xi_r \end{pmatrix} \geq 0.$$

Since $x-\xi_1$, $x-\xi_2$, ..., $x-\xi_r$ are in decreasing order of magnitude, it follows from Theorem 3 that

$$h\begin{pmatrix} n_1, & n_2, \cdots, & n_r \\ x-\xi_1, x-\xi_2, \cdots, x-\xi_r \end{pmatrix}$$

has the sign $(-1)^{r(r-1)/2}$. Thus

$$c_+\begin{pmatrix} n_1, n_2, \cdots, n_r \\ m_1, m_2, \cdots, m_r \end{pmatrix}$$

has the sign $(-1)^{r(r-1)/2}$, as was to be proved.

We prove below in Theorem 3 that c(n) is PF_2 provided f(x) is PF_2. This is the property required for the analysis of the inventory model of Section 2. In contrast, this relationship between c(n) and f(x) does not persist beyond the second order.

THEOREM 3: If f(x) is PF_2 with f(x) = 0 for $x < 0$, then c(n) (defined in Theorem 2) is PF_2.

PROOF: Let $n_1 < n_2$, $m_1 < m_2$. Write

$$c\begin{pmatrix} n_1, & n_2 \\ m_1, & m_2 \end{pmatrix} = \begin{vmatrix} c(n_1 - m_1) & c(n_1 - m_2) \\ c(n_2 - m_1) & c(n_2 - m_2) \end{vmatrix}.$$

(a) If $n_1 \leq m_2$, then $c(n_1 - m_2) = 0$, so that $c\begin{pmatrix} n_1, & n_2 \\ m_1, & m_2 \end{pmatrix} \geq 0$.

(b) If $n_1 > m_2$, we must have $m_1 < m_2 < n_1 < n_2$. Hence

$$c\begin{pmatrix} n_1, & n_2 \\ m_1, & m_2 \end{pmatrix} = \begin{vmatrix} h(n_1 - m_1, x) & h(n_1 - m_2, x) \\ h(n_2 - m_1, x) & h(n_2 - m_2, x) \end{vmatrix}$$

$$= \int \begin{vmatrix} h(n_1 - m_2, \xi) & h(n_1 - m_2, x) \\ h(n_2 - m_2, \xi) & h(n_2 - m_2, x) \end{vmatrix} g(m_2 - m_1, x-\xi)\,d\xi .$$

Since $\xi < x$, $n_1 - m_2 < n_2 - m_2$, and $h(n, x)$ is TP_2 by Theorem 1 of this section, then $c\binom{n_1, n_2}{m_1, m_2} \geq 0$, and the proof is finished.

BIBLIOGRAPHY

[1] Arrow, K., Karlin, S., and Scarf, H., Studies in the Mathematical Theory of Inventory and Production, Stanford University Press, 1958.

[2] Black, G. and Proschan, F., "On Optimal Redundancy," Operations Research, Vol. 7, No. 5, Sept. - Oct. 1959.

[3] Black, G. and Proschan, F., "Spare Parts At Minimum Cost," Proceedings of the Fifth Nat'l Symp. on Reliability and Quality Control, 1959, originally appearing as Electronic Defense Laboratory Technical Memorandum EDL-M154, 13 Oct. 1958.

[4] Cramer, H., Mathematical Methods of Statistics, Princeton University Press, 1946.

[5] Gantmacher, F. and Krein, M., Oscillation Matrices and Kernels and Small Vibrations of Mechanical Systems, 2nd ed. (in Russian), Moscow, 1950.

[6] Geisler, M. A. and Karr, H. W., "The Design of Military Supply Tables for Spare Parts," Operations Research, Vol. 4, No. 4, August 1956, pp. 431-442.

[7] Gourary, M. H., "A Simple Rule for the Consolidation of Allowance Lists," Naval Research Logistics Quarterly, Vol. 5, No. 1, March 1958, pp. 1-15.

[8] Gourary, M. H., "An Optimum Allowance List Model," Naval Research Logistics Quarterly, Vol. 3, 1956, pp. 177-191.

[9] Hirschmann, I. I. and Widder, D. V., The Convolution Transform, Princeton University Press, Princeton, New Jersey, 1955.

[10] Jorgensen, J., "Optimal Spares Provisioning for EDL Systems," Electronic Defense Laboratory Report EDL-M237, 15 Jan. 1960.

[11] Karlin, S., Mathematical Methods and Theory In Games, Programming, and Economics, Addison Wesley (1960).

[12] Karlin, S., "Polya Type Distributions IV: Some Principles of Selecting a Single Procedure from a Complete Class," Annals of Mathematical Statistics, Vol. 29, No. 1, March 1958, pp. 1-21.

[13] Karlin, S., "Polya Type Distributions III: Admissibility of Multiaction Problems," Annals of Mathematical Statistics, Vol. 28, No. 4, December 1957, pp. 839-860.

[14] Karlin, S., "Polya Type Distributions II," Annals of Mathematical Statistics, Vol. 28, No. 2, June 1957, pp. 281-308.

[15] Karlin, S., "Decision Theory for Polya Type Distributions. Case of Two Actions, I," Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, December 1954, and June and July 1955, pp. 115-128.

[16] Karlin, S., and McGregor, J., "Coincident Properties of Birth and Death Processes," Applied Mathematics and Statistics Laboratories, Stanford University, Technical Report No. 9, Under Contract Nonr-225(28) for Office of Naval Research, Nov. 18 1958.

[17] Karlin, S., and Proschan, F., "Polya Type Distributions of Convolutions," Applied Mathematics and Statistics Laboratory, Stanford University, Technical Report No. 14, prepared under Contract Nonr-225(28) for Office of Naval Research, August 14 1959; submitted to Annals of Mathematical Statistics.

[18] Karlin, S., and Rubin, H., "The Theory of Decision Procedures for Distributions with Monotone Likelihood Ratio," Annals of Mathematical Statistics, Vol. 27 1956.

[19] Kuhn, H. W., and Tucker, A. W., "Nonlinear Programming," Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1951, pp. 481-492.

[20] Molina, E. C., "Poisson's Exponential Binomial Limit," D. Van Nostrand Co., 1942.

[21] Polya, G., and Szego, G., Ausgaben und Lehrsatze Aus Der Analyse, Vol. 1, Berlin, Julius Springer, 1925.

[22] Proschan, F., "Periodic Resupply Policy Achieving Specified Protection Against System Shutdown At Minimum Cost," Electronic Defense Laboratory Technical Memorandum No. EDL-M205, 12 June 1959.

[23] Proschan, F., "Minimum Cost for Spares When Component Failure Is Not Necessarily Exponential," Electronic Defense Laboratory Technical Memorandum No. EDL-M185, 14 April 1959.

[24] Radio Corporation of America, "Reliability Stress Analysis For Electronic Equipment," Components Application Review Monograph No. 3, Technical Report TR 1100, 28 November 1956.

[25] Schoenberg, I. J., "On Polya Frequency Functions. I. The Totally Positive Functions and Their Laplace Transforms," Journal D'Analyse Mathematique, Vol. I, 1951, pp. 331-374.

* * *

INVESTMENT POLICIES FOR EXPANDING BUSINESSES OPTIMAL IN A LONG-RUN SENSE

Leo Breiman

University of California

An entrepeneur has a given initial fortune and faces the following situation: during any time period he may invest various amounts of his available fortune in various alternatives (differing from interval to interval) and keep the remainder of his money as a reserve fund. The alternatives materialize and pay off according to a probability distribution in which occurrences in different time intervals are independent. We analyze the problem of finding the investment policies (that is, the division of funds during every time interval between the investment alternatives and the reserve) that will make the long-run growth of the entrepeneur's fortune as rapid as possible.

INTRODUCTION AND SUMMARY

The situation we treat is as follows: at the beginning of the N^{th} investment period, an entrepeneur has fortune S_N and is prescribed with investment alternatives $1, 2, \ldots, n$. Associated with these alternatives is a random pay-off vector $\overline{r}_N = (r_1, \ldots, r_n)$, $r_i \geq 0$, having some distribution in n-dimensional space, so that if the investor puts amounts $S_i, \ldots, S_n$ of his fortune into alternatives $1, \ldots, n$, then his return at the end of the investment period from alternative i is $r_i\ S_i$. The question is whether there is any optimal investment policy, that is, any optimal way of dividing the available investment fund between the various alternatives. (We note that one of the alternatives could be to put so much money into a reserve fund which then draws a given interest rate.)

Of course, the optimal policies will vary with our criterion of optimality. One classical criterion is to maximize our expected return; that is, invest $S_i, \ldots, S_n, \sum_1^n S_r = S_N$ such that $E\left(\sum_1^n S_i\ r_i\right)$ = maximum. This criterion leads to the highly speculative policy: invest all of S_N in that alternative i such that $Er_i = \max E\ r_j$. A policy of this type is clearly untenable to any continuing investor or firm, since it usually leads to quick bankruptcy proceedings. For example, the alternative with maximum expected return may have $r = 0$ with high probability and r large with small probability. A repeated gambling of all available funds on this alternative will, with probability one, lead to ultimate ruin. One might say that, in this situation, the right thing to do during every investment period is to hold part of one's capital in reserve and invest the remainder in the alternative with the maximum expected return. The obvious objection to this is that the investor is no longer following his optimum policy; besides, what criterion is being used to decide what funds to hold in reserve?

We have discussed the undesirability of the above criterion from the point of view of ultimate ruin; that is, from the long-run point of view. In our opinion, most non-speculative investment policies implicitly take this view. In order to formalize this criterion, we assume that we start with an initial fortune S_0, and that S_{N+1}, our fortune at the beginning of the $N+1^{st}$ period, consists of the total returns from our investments of the N^{th} period; i.e. everything that we make during the N^{th} period is plowed back into investments during the $N+1^{st}$ period. This assumption is made in order to be able to compare the asymptotic behavior of S_N under different investment policies, and is approximately valid during the initial expansion period of business firms and in investment holdings where the dividends taken out of the system are small compared with the total dividends. Any well-formulated rule for the removal of funds from the system will evidently enable an asymptotic analysis of S_N; we use the above rule, however, for simplicity and illustration.

Concerning the investment alternatives, we assume very little; they may vary arbitrarily from period to period and the distribution of the pay-off vector r_N for the N^{th} period may be conditioned in any way by the past. Perhaps our most drastic assumption is that any division of funds between the various alternatives is an allowable investment policy. This is unrealistic in the many situations where there is either an upper limit to investment, or a lower limit, or both.

Naturally, our investment policy for the N^{th} period may depend both on S_N and on the outcomes of the past. Now let S_N be the fortune of the beginning of the N^{th} period under one sequence of investment policies and S'_N the fortune of the same time under a competing sequence of investment policies. We will say that the former sequence is inadmissible if there is a fixed number $\alpha > 0$ such that for every $\epsilon > 0$ there exists a competing sequence such that on a set of probability greater than $\alpha - \epsilon$

$$\lim_N \sup \frac{S_N}{S'_N} = 0$$

and, except for a set of probability at most ϵ,

$$\lim_N \sup \frac{S_N}{S'_N} \leq 1 .$$

Roughly, the sequence leading to S_N is inadmissible if for every $\epsilon > 0$ there is a competing sequence that is asymptotically infinitely better with probability at least $\alpha-\epsilon$ and with, at most, probability ϵ of being worse.

Restricting ourselves to admissible sequences of policies, we domonstrate that there is a unique admissible sequence which is distinguished by the fact that every admissible sequence must be asymptotically close to it. This sequence is found as follows: let $\lambda_1 \ldots, \lambda_n$ be the proportions of S_N that we invest in alternatives 1, ..., N, then choose the λ_i so as to maximize

$$E \log \left(\sum_1^n \lambda_i r_i \right) , \quad \lambda_i > 0, \ \sum_1^n \lambda_i = 1 .$$

It is easily shown that this is equivalent to choosing the λ_i so that they satisfy

$$\lambda_i = E\left(\frac{\lambda_i r_i}{\sum_i \lambda_i r_i}\right).$$

Hence, we may verbally state this investment policy as: let the investment proportion in alternative i be equal to the expected proportion of the return coming from alternative i. It is interesting, as well as somewhat gratifying, that this policy is one of diversification; that is, in general, this policy is one of dividing S_N between a number of alternatives.

Finally, as far as the mathematics of the situation is concerned, our main tool is the ever-reliable martingale theorem.

MATHEMATICAL FORMULATION AND PROOFS

Let $\bar{\lambda}_N = (\lambda_1, \ldots, \lambda_n)$ be the investment proportion vector for the N^{th} period, i.e. λ_i is the proportion of S_N invested in alternative i; so using the notation $V_N = \sum_1^n \lambda_i r_i$ we have

$$S_{N+1} = S_N V_N .$$

Now let R_{N-1} be the outcomes during the first N-1 investment periods, $R_{N-1} = (\bar{r}_{N-1}, \ldots, \bar{r}_1)$. We consider two competing sequences of investment policies $(\bar{\lambda}_1, \bar{\lambda}_2, \ldots)$ and $(\bar{\lambda}_1^*, \bar{\lambda}_2^*, \ldots)$ such that under each we start with the same initial fortune S_0. Now λ_N, λ_N^* may depend in an arbitrary manner on R_{N-1} as may the distribution of r_N. Let S_N be the fortune under $(\bar{\lambda}_1, \ldots)$ and S_N^* the fortune under $(\bar{\lambda}_1^*, \ldots)$ where we take $\bar{\lambda}_N^*$ to be defined as that investment proportion vector which maximizes

$$E(\log V_N \mid R_{N-1}), \qquad V_N = \sum_i \lambda_i r_i, \ \lambda_i \geq 0, \sum_i \lambda_i = 1 .$$

We need some uniform integrability condition, and we will assume that there are constants α, β such that

$$0 < \alpha < r_i < \beta < \infty , \text{ all } i, N.$$

The theorem is true under much weaker conditions, but with the above condition, it reduces to a standard martingale theorem and thus keeps the technicalities at a minimum.

THEOREM 1: Under (A), limit of S_N/S_N^* exists almost surely and $E(\lim S_N/S_N^*) \leq 1$.

PROOF: We have

(A) $$E\left(\frac{S_N}{S_N^*} \mid R_{N-1}\right) = E\left(\frac{V_N}{V_N^*} \mid R_{N-1}\right) \frac{S_{N-1}}{S_{N-1}^*} .$$

Hence, it is sufficient to prove that $E\left(V_N/V_N^* \mid R_{N-1}\right) \leq 1$, for then S_N/S_N^* is a decreasing semimartingale with $E(\lim S_N/S_N^*) \leq E(S_0/S_0^*) = 1$ (see [1]). Now, for any $\varepsilon > 0$, by the maximizing property of $\bar{\lambda}_N^*$ and by (A), we have

$$E\left\{\log\left[(1-\varepsilon)V_N^* + \varepsilon V_N\right] - \log V_N^* \mid R_{N-1}\right\} < 0,$$

so that

$$\frac{1}{\varepsilon} E\left[\log\left(1 + \frac{\varepsilon}{1-\varepsilon}\frac{V_N}{V_N^*}\right)\Big| R_{N-1}\right] < \frac{1}{\varepsilon}\log\frac{1}{1-\varepsilon}.$$

By (A), V_N/V_N^* is bounded, and we let $\varepsilon \to 0$ to get the desired result.

THEOREM 2: The set on which $\lim_N S_N/S_N^* = 0$ is almost surely equal to the set on which

$$\sum_{N=1}^{\infty}\left[E\left(\log V_N^* \mid R_{N-1}\right) - E\left(\log V_N \mid R_{N-1}\right)\right] = \infty.$$

PROOF: The sequence

$$\log\frac{S_N}{S_N^*} - E\left(\log\frac{S_N}{S_N^*}\Big| R_{N-1}\right) = \sum_{k=1}^{N}\left[\log\frac{V_k}{V_k^*} - E\left(\log\frac{V_k}{V_k^*}\Big| R_{k-1}\right)\right]$$

is a martingale sequence with

$$\sup_{k \geq 1}\left|\log\frac{V_k}{V_k^*} - E\left(\log\frac{V_k}{V_k^*}\Big| R_{k-1}\right)\right| \leq \gamma < \infty.$$

Therefore (see [1], pp. 319-320), the sequence converges to a finite value almost surely, which yields theorem 3.

THEOREM 3: Under (A) the sequence $(\lambda_1, \lambda_2, \ldots)$ is admissible if, and only if, almost surely,

$$\text{(B)} \qquad \sum_{1}^{\infty}\left[E(\log V_k^* \mid R_{k-1}) - E(\log V_k \mid R_{k-1})\right] < \infty.$$

PROOF: Suppose (B) is violated, then there is a set E of positive probability (i.e. $P(E) = \alpha > 0$) such that on E, $\lim_N S_N/S_N^* = 0$. For any $\varepsilon > 0$, there exists a set E_M, measurable with respect to R_M such that $P(E_M \Delta E) < \varepsilon$, where Δ denotes the symmetric set difference. We define a competing sequence of policies $(\bar{\lambda}_1', \bar{\lambda}_2', \ldots)$ as follows: if

$N-1 < M$, $\bar{\lambda}_N = \bar{\lambda}_N$, if R_{N-1}, $N-1 \geq M$, is such that $(\bar{r}_1, \ldots, \bar{r}_M)$ is not in E_M, then $\bar{\lambda}'_N = \bar{\lambda}_N$, otherwise $\bar{\lambda}'_N = \bar{\lambda}^*_N$. Now, on E_M,

$$\sum_1^\infty \left[E(\log V_k^*|R_{k-1}) - E(\log V_k'|R_{k-1})\right] < \infty$$

hence, $\lim S'_N/S^*_N > 0$ which implies that $\lim S_N/S'_N = 0$ on $E_M \cup E$. And we have that $P(E_M \cup E) > P(E) - \varepsilon = \alpha - \varepsilon$, on the complement of E_M, $S_N = S'_N$, so that $\lim S_N/S'_N \leq 1$ except with at most probability ε. Therefore, $(\bar{\lambda}_1, \bar{\lambda}_2 \ldots)$ is an inadmissible sequence.

Conversely, suppose (B) holds, but $(\bar{\lambda}_1, \ldots)$ is inadmissible with respect to $(\bar{\lambda}'_1 \ldots)$, and that $\lim S_N/S'_N = 0$ on E with $P(E) > 0$. Since (B) holds, $\lim S_N/S^*_N > 0$ almost surely, which implies that $\lim S^*_N/S'_N = 0$ almost surely on E. This implies, in turn, that $\lim S'_N/S^*_N = \infty$ with positive probability, which violates Theorem 1.

The above Thoerem is the result referred to in the introduction, for the essential content is, that unless $\bar{\lambda}_N$ is usually close to $\bar{\lambda}^*_N$ in the sense that $E(\log V^*_N|R_{N-1})$ $- E(\log V_N|R_{N-1})$ is small, then $(\bar{\lambda}_1, \ldots)$ is not admissible.

REFERENCE

[1] J. L. Doob, Stochastic Processes, John Wiley and Sons, New York, 1953.

* * *

NOTES

PROBLEMS

It has been suggested that this journal might serve as a medium for publishing problems in the area of Logistics — with the idea that other persons might be interested in those problems and might submit comments. Readers are invited to submit brief statements on applied and theoretical problems in Logistics. Address letters to Managing Editor, Naval Research Logistics Quarterly, Office of Naval Research, Washington 25, D. C.

- - - - - - - - - - -

ERRATA

In the December 1959 issue of NRLQ (Vol. 6, No. 4), p. 335, first line - Stroller, D.S. should be listed Stoller, D. S.

NEWS AND MEMORANDA

Readers Are Invited To Submit To The Managing Editor Items Of General Interest In The Field Of Logistics

Leon Gilford has accepted a position as Mathematical Statistician with Operations Research Incorporated, Silver Spring, Maryland. He was formerly Chief of the Operations Research Branch, Statistical Research Division, U. S. Bureau of the Census.

AUTHOR INDEX FOR VOLUME 7

AUMANN, R. J. - "Linearity of Unrestrictedly Transferable Utilities," Vol. 7, No. 3, Sept. 1960, pp. 298 - 291

BREIMAN, Leo - "Investment Policies for Expanding Businesses Optimal in a Long-Run Sense," Vol. 7, No. 4, Panel 6 - Inventory Control

BRODSKY, Nathan - "The Need for a Strategy of Logistics Research" Vol. 7, No. 4, Introductory Remarks

BROWN, Bernice - "A Compariative Study of Prediction Techniques," Vol. 7, No. 4, Panel 3b - Mico-Simulation in Logistics Research

CANNON, E. W. - "The Reflection of Logistics in Electronic Computer Design," Vol. 7, No. 4, Panel 1 - Hardware for Logistics Research

CHARNES, A., COOPER, W. W. - "Some Problems and Models for Time-Phased Transport Requirements," Vol. 7, No. 4, Panel 5 - Transportation and Scheduling Problems

CLARK, A. J. - "The Use of Simulation to Evaluate a Multiechelon, Dynamic Inventory Model," Vol. 7, No. 4, Panel 3a - Macro-Simulation in Logistics Research

DAVID, Harold - "The Build-Up Time of Waiting Lines," Vol. 7, No. 2, June 1960, pp. 185 - 193

DENICOFF, M., FENNELL, J. P., SOLOMON, H. - "Summary of a Method for Determining the Military Worth of Spare Parts," Vol. 7, No. 3, Sept. 1960, pp. 221 - 234

DERMAN, C., J. SACKS - "Replacement of Periodically Inspected Equipment," Vol. 7, No. 4, Panel 6 - Inventory Control

D'ESOPO, A., DIXION, H. L., LEFKOWITZ, B., - "A Model for Simulating an Air-Transportation System," Vol. 7, No. 3, Sept. 1960, pp. 213 - 220

DRESCH, Francis - "On-Line Macro-Simulation in Systems for Logistics Decision Making," Vol. 7, No. 4, Panel 3a - Macro-Simulation in Logistics Research

FENNELL, J. P. - "An Automatic Addressing Device," Vol. 7, No. 4, Panel 1 - Hardware for Logistics Research

FIRSTMAN, Sidney I. - "An Approximation Algorithm for an Optimum Aim-Points Problem," Vol. 7, No. 2, June 1960, pp. 151 - 167

FOLSOM, P. L. - "Military Worth and Systems Development," Vol. 7, No. 4, Panel 4 - Utility and Military Worth

FRANK, P. - "On Assuring Safety in Destructive Testing," Vol. 7, No. 3, Sept. 1960, pp. 257 - 259

MANNE, A. S. - "Comments on 'Interindustry Economics' by Chenery and Clark," Vol. 7, No. 4, Dec. 1960, Panel 2 - Numerical Description of the Economy

MARLOW, W. H. - "Some Accomplishments of Logistics Research," Vol. 7, No. 4, Dec. 1960, Plenary Session

MELLON, W. G. - "Priority Ratings in More Than One Dimension," Vol. 7, No. 4, Dec. 1960, Panel 4 - Utility and Military Worth

NOBLE, S. B. - "Some Flow Models of Production Constraints," Vol. 7, No. 4, Dec. 1960, Panel 2 - Numerical Description of the Economy

O'NEILL, R. R. - "Scheduling of Cargo Containers," Vol. 7, No. 4, Panel 5 Transportation and Scheduling Problems

OKUN, Bernard - "Design, Test, and Evaluation of an Experimental Flyaway Kit," Vol. 7, No. 2, June 1960, pp. 109 - 136

ORCHARD-HAYS, William - "Control of and Communication with Data-Handling Machines," Vol. 7, No. 4, Dec. 1960, Panel 1 - Hardware for Logistics Research

PROSCHAN, Frank - "Optimal System Supply," Vol. 7, No. 2, Dec. 1960, Panel 6 - Inventory Control

SCARF, H. E. - "Some Remarks on Bayes Solutions to the Inventory Problem," Vol. 7, No. 4, Dec. 1960, Panel 6 - Inventory Control

SCOFIELD, E. K. - "Research for Command Logistics," Vol. 7, No. 4, Dec. 1960, Plenary Session

SOLOMON, Henry - "The Determination and Use of Military-Worth Measurements for Inventory Systems," Vol. 7, No. 4, Dec. 1960, Panel 4, Utility and Military Worth

SOLOMON, Henry, DENICOFF, Marvin - "Simulations of Alternative Allowance List Policies," Vol. 7, No. 2, June 1960, pp. 137 - 149

STEINHAUS, H. - "Definitions for a Theory of Games and Pursuit," Vol. 7, No. 2, June 1960, pp. 105 - 108

TEAGER, H. M. - "The Marriage of On-Line Human Decision with Computer Programs," Vol. 7, No. 4, Dec. 1960, Panel 1 - Hardware for Logistics Research

THOMPSON, G. L. - "Recent Developments in the Job-Shop Scheduling Problem," Vol. 7, No. 4, Dec. 1960, Panel 5 - Transportation and Scheduling Problem

TOMPKINS, C. B. - "Some Aspects of Mathematics in Social Sciences," Vol. 7, No. 4, Dec. 1960, Plenary Session

GARFUNKEL, I. M., WALSH, John E. - "Method for First-Stage Evaluation of Complex Man-Machine Systems," Vol. 7, No. 1, March 1960, pp. 13 - 19

GEISLER, M. A. - "A First Experiment in Logistics System Simulation," Vol. 7, No. 1, March 1960, pp. 21 - 44

GEISLER, M. A. - "The Use of Man-Machine Simulation for Support Planning" Vol. 7, No. 4, Dec. 1960, Panel 3a - Macro-Simulation in Logistics Research

GIFFLER, Bernard - "Determining and Optimum Reject Allowance," Vol. 7, No. 2, June 1960, pp. 201 - 206

GLICKSMAN, Stephen, JOHNSON, Lyle, ESELSON, Leonard, - "Coding the Transportation Problem," Vol. 7, No. 2., June 1960, pp. 169 - 183

GLUSS, Brian - "An Optimal Inventory Solution for Some Specific Demand Distribution," Vol. 7, No. 1, March 1960, pp. 45 - 48

HAMILTON, J. E. - "Realism in Empirical Logistics Research," Vol. 7, No. 4, Dec. 1960, Panel 3b - Micro Simulation in Logistics Research

HARARY, Frank - "A Matrix Criterion for Structural Balance," Vol. 7, No. 2, June 1960, pp. 195 - 199

HUNT, R. B., ROSHOLDT, E. F. - "Determining Merchant Shipping Requirements in Integrated Military Planning," Vol. 7, No. 4, Dec. 1960, Panel 5 - Transportation and Scheduling Problems

HUNT, Ralph B., ROSHOLDT, Erling F. - "The Concepts of Notional Ship and Notional Value in Logistics Capability Studies Involving Merchant Ships," Vol. 7, No. 1, March 1960, pp. 1 - 6

JACKSON, J. R. - "Some Problems in Queueing with Dynamic Priorties," Vol. 7, No. 3, Sept. 1960, pp. 235 - 249

KARREMAN, H. F. - "Programming the Supply of a Strategic Material - Part I A Non-stochastic Model," Vol. 7, No. 3, Sept. 1960, pp. 261 - 279

KENDRICK, John - "Changing Output-Input Relations within the National Economic Accounts," Vol. 7, No. 4, Dec. 1960, Panel 2 - Numerical Description of the Economy

KLEIN, M., ROSENBERG, L. - "Deterioration of Inventory and Equipment," Vol. 7, No. 1, March 1960, pp. 49 - 62

WAGNER, Harvey M. - "A Postscript to 'Dynamic Problems in the Theory of the Firm,'" Vol. 7, No. 1, March 1960 pp. 7 - 12

WALSH, J. E. - "A General Simulation Model for Logistics Operation in a Randomly Damaged System," Vol. 7, No. 4, Dec. 1960, Panel 3a - Macro-Simulation in Logistics Research

WEISS, L. - "Confidence Intervals of Preassigned Length for Qualities of Unimodal Populations," Vol. 7, No. 3, Sept. 1960, pp. 251 - 256

* U.S. GOVERNMENT PRINTING OFFICE : 1961 O—600250

CONTENTS (Continued)

Changing Output-Input Relations Within the National Economic Accounts 393
J. W. Kendrick

Some Flow Models of Production Constraints 401
S. B. Noble

PANEL 3a - MACROSIMULATION IN LOGISTICS RESEARCH

The Use of Man-Machine Simulation for Support Planning 421
M. A. Geisler

The Use of Simulation to Evaluate a Multiechelon, Dynamic Inventory Model 429
A. J. Clark

On-Line Macro-Simulation in Systems for Logistics Decision Making 447
Francis Dresch

General Simulation Model for Logistics Operation in a Randomly Damaged System 453
J. E. Walsh

PANEL 3b - MICROSIMULATION IN LOGISTICS RESEARCH

A Comparative Study of Prediction Techniques 471
Bernice Brown

Realism in Empirical Logistics Research 493
J. E. Hamilton

PANEL 4 - UTILITY AND MILITARY WORTH

Military Worth and Systems Development 501
P. L. Folsom, CAPT., USN

Priority Ratings in More Than One Dimension 513
W. G. Hellon

The Determination and Use of Military-Worth Measurements for Inventory Systems 529
Henry Solomon

PANEL 5 - TRANSPORTATION AND SCHEDULING PROBLEMS

Some Problems and Models for Time-Phased Transport Requirements 533
A. Charnes and W. W. Cooper

Determining Merchant Shipping Requirements in Integrated Military Planning 545
R. B. Hunt and E. F. Rosholdt

Scheduling of Cargo Containers 577
R. R. O'Neill

Recent Developments in the Job-Shop Scheduling Problem 585
G. L. Thompson

PANEL 6 - INVENTORY CONTROL

Some Remarks on Bayes' Solutions to the Inventory Problem 591
H. E. Scarf

Replacement of Periodically Inspected Equipment 597
C. Derman and J. Sacks

Optimal System Supply 609
Frank Proschan

Investment Policies for Expanding Businesses Optimal in a Long-Run Sense 647
Leo Breiman

NOTES 653

NEWS AND MEMORANDA 655

AUTHOR INDEX FOR VOLUME 7 657

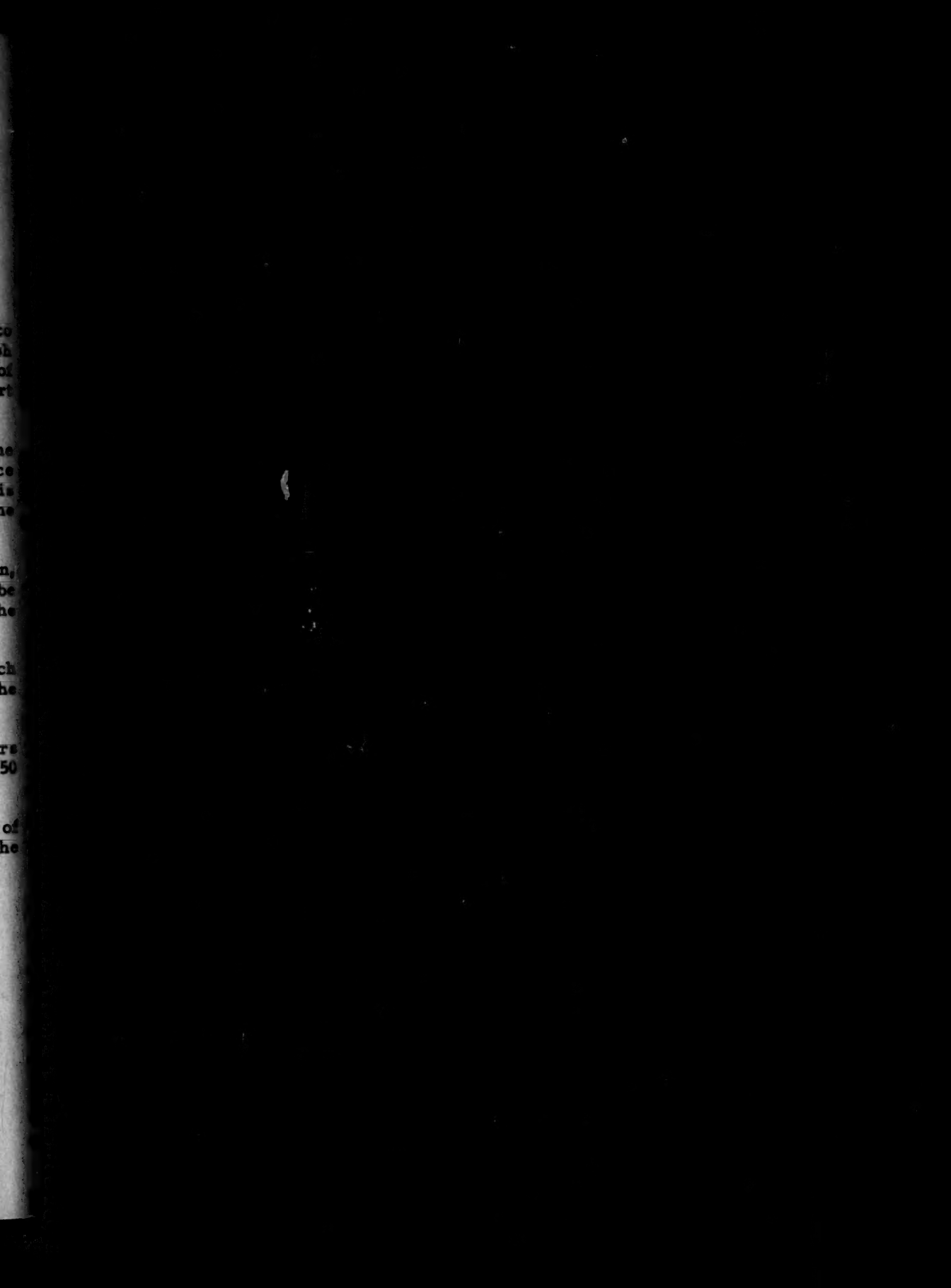
INFORMATION FOR CONTRIBUTORS

The NAVAL RESEARCH LOGISTICS QUARTERLY is devoted to the dissemination of scientific information in logistics and will publish research and expository papers, including those in certain areas of mathematics, statistics, and economics, relevant to the over-all effort to improve the efficiency and effectiveness of logistics operations.

Manuscripts and other items for publication should be sent to The Managing Editor, NAVAL RESEARCH LOGISTICS QUARTERLY, Office of Naval Research, Washington 25, D. C. Each manuscript which is considered to be suitable material for the QUARTERLY is sent to one or more referees.

Manuscripts submitted for publication should be typewritten, double-spaced, and the author should retain a copy. Refereeing may be expedited if an extra copy of the manuscript is submitted with the original.

A short abstract (not over 400 words) should accompany each manuscript. This will appear at the head of the published paper in the QUARTERLY.

There is no authorization for compensation to authors for papers which have been accepted for publication. Authors will receive 50 reprints of their published papers.

Readers are invited to submit to the Managing Editor items of general interest in the field of logistics, for possible publication in the NEWS AND MEMORANDA or NOTES sections of the QUARTERLY.